DOCUMENT RESUME

ED 212 497 SE 036 246

AUTHOR Tatsuoka, Kikumi K.; Tatsuoka, Maurice M.
TITLE Item Analysis of Tests Designed for Diagnosing Bugs: Item Relational Structure Analysis Method.
INSTITUTION Illinois Univ., Urbana. Computer-Based Education Research Lab.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
REPORT NO CERL-RR-81-7.
PUB DATE Nov 81
GRANT NIE-G-81-0002
NOTE 27p.; For related document, see SE 036 245.
AVAILABLE FROM Kikumi Tatsuoka, Computer-Based Education Research Lab., 252 Engineering Research Lab., 103 S. Mathews, Univ. of Illinois at Urbana, Urbana, IL 61801 (no price quoted).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Addition; *Computation; Educational Research; Elementary Secondary Education; Evaluation; *Evaluation Methods; *Fractions; *Item Analysis; *Mathematics Education; Models; Problem Solving; Subtraction; Testing; *Test Items
IDENTIFIERS *Mathematics Education Research

ABSTRACT

A new system of order analysis, developed by Takeya and called Item Relations Structure Analysis (IRSA), was described and used for examining the structural relations among a set of 24 items on the addition and subtraction of fractions. A diagraph showing 16 chains of items that had discernibly common features was generated by this method, and implications for diagnostic error analysis were discussed. (Author)

CERL Computer-based Education Research Laboratory

University of Illinois Urbana Illinois

ITEM ANALYSIS OF TESTS DESIGNED FOR DIAGNOSING BUGS: ITEM RELATIONAL STRUCTURE ANALYSIS METHOD

KIKUMI K. TATSUOKA
MAURICE M. TATSUOKA

This research was partially supported by the National Institute of Education, under the grant No. NIE-G-81-0002. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by the National Institute of Education should be inferred.

COMPUTERIZED ADAPTIVE TESTING AND MEASUREMENT RESEARCH REPORT 81-7
NOVEMBER 1981

Copies of this report may be requested from

Kikumi Tatsuoka
Computer-based Education Research Laboratory
252 Engineering Research Laboratory
103 S. Mathews
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Abstract

A new system of order analysis, developed by Takeya and called Item Relations Structure Analysis (IRSA), was described and used for examining the structural relations among a set of 24 items on the addition and subtraction of fractions. A digraph showing 16 chains of items that had discernibly common features was generated by this method, and implications for diagnostic error analysis were discussed.

Acknowledgment

The authors wish to acknowledge the kind cooperation extended to us by the people involved with this report. Bob Baillie programmed the data collection and analysis routines, along with his assistant, David Dennis. Mary Klein administered the tests to her junior high school students. Roy Lipschutz did the layouts and Louise Brodie did the typing.

Introduction

Tatsuoka & Tatsuoka (1981) demonstrated in their study that even a criterion-referenced test, in which items are chosen from a single content domain, violates a homogeneity of the test items when students use a variety of different methods to solve the problems. Thus, examining the underlying cognitive processes that are adopted by the students is very important before any test theory models, such as Item Response Theory or criterion-referenced test theory, are applied for analyzing the performances of the students on the tests. These modern test theories require some strict assumptions on the structure of an item domain, although they are very useful in many ways.

Investigating the structure of test items can be done by several different procedures -- factor analysis, scalogram analysis (Guttman, 1950) Loevinger's (1948) measure of test homogeneity and order analysis (Krus, 1975, Cliff, 1977). Unfortunately, these procedures have failed to produce satisfactory results from achievement data obtained from a series of experimental studies (Tatsuoka & Birenbaum, 1979, 1981; Birenbaum & Tatsuoka, 1981; Tatsuoka & Tatsuoka, 1980).

Order analysis has been used in constructing a hierarchical structure of the items and instructional units (Airasian & Bart, 1973; Bart & Krus, 1973). Wise (1981) has developed a new order-analysis procedure to extract unidimensional subsets from the total set of test items and Takeya (1981) has defined a new order structure by using the expected proportions of dominance relationships between two items. Takeya's order structure is mathematically elegant, and it has algebraic relations with Loevinger's homogeneity index, Mokken's index (Mokken, 1971), caution index (Sato, 1975) and Cliff's index C_{t3}.

Therefore, we will adopt Takeya's order analysis (called Item Relations Structure Analysis, IRSA) to examine the item relationship of fraction problems (Klein, et al., 1981). The advantage of using IRSA is (according to Takeya) that it enables us to see a cognitive aspect of the students' performances on the items to a certain extent. Since it generates a digraph representing the hierarchcal structure of the items, it will -- at the very least -- allow us to check the extent to which we have succeeded in constructing problems that require a hierarchically specified set of skills for solving them. Thus, with some refinement and further development, it should enable us to construct tests for diagnostic purposes with built-in remedial tasks.

A Brief Summary of the Theoretical Background of Order Analysis

In the traditional scoring of an n-item test, each item receives a score of 1 when the response to the item matches the key and 0 when it does not. For each subject k the resulting response pattern can be represented by a row vector:

$$\underline{x}_k = (x_{k1}, \ldots, x_{kj}, \ldots, x_{kn});$$
$$x_{kj} = 1, 0;\ k=1\ldots, N;\ j=1,\ldots n.$$

Mokken (1971) defines "a perfect scale" by describing it as a step function: $P(x_{kj} = 1|\theta) = 0$ if subject k's ability level θ_k is located to the left of the difficulty level, d_j of item j on the horizontal axis and 1 if the θ_k is to the right of d_j. The following figure illustrates the perfect-scale function.

Insert Figure 1 about here

If all subjects perform on item j according to the Mokken's perfect-scale function, then item j is said to be a perfect item. However, real data usually don't follow this logic and quite a number of subjects score 0 even if $\theta_k > d_j$. A contingency table of two items, item i with difficulty d_i less than that of another item, d_j, will explain the situation. Let the items be arranged in ascending order of difficulty, so that $d_i \leq d_j$ when $i < j$. (Lower-numbered items are easier than higher-numbered items.)

		item j: 1	item j: 0	
item i	1	n_{11}	n_{10}	n_i
	0	n_{01}	n_{00}	$N-n_i$
		n_j	$N-n_j$	N

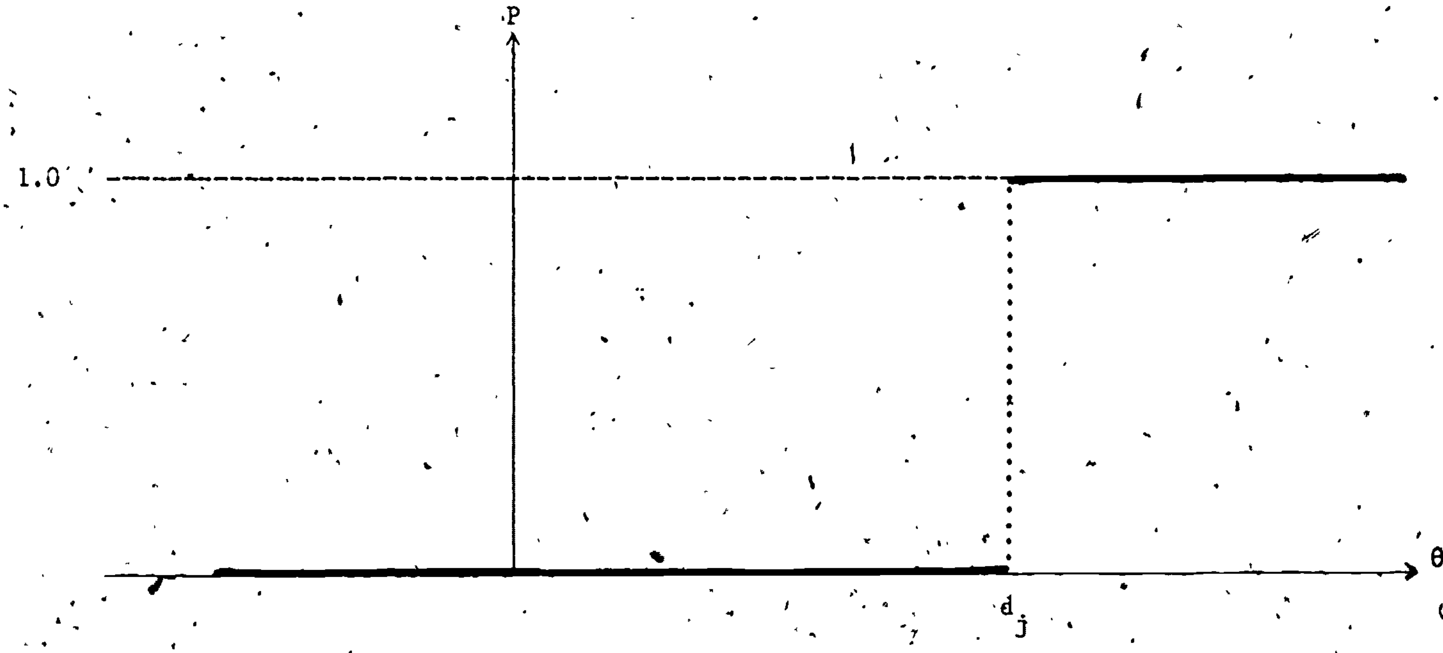

Figure 1. Mokken's Perfect-Scale Function: $P(x_{kj} = 1|\theta) = 0$ if $\theta < d_j$

$= 1$ if $\theta \geq d_j$

If items i and j are perfect then $n_{01} = 0$. We call the response pattern (0,1), with item i wrong but item j right, "inconsistent." For a set of N subjects, if the inconsistent cell (0,1) is empty (i.e., if $n_{01} = 0$) then the set of N two-dimensional response vectors forms a Guttman scale.

A number of researchers have tried to cope with the problem of inconsistent responses and have developed various types of measures (or used pre-existing ones) -- the proximity of two items, agreement coefficient, ϕ coefficient, eta coefficient, tetrachoric correlation, etc. These measures express some aspects of the relationship between the two items, but not all aspects.

Takeya's index r^*_{ij} is based on the dominance relation between items i and j and considers the statistical independence or dependence of the scores obtained from the two items. Now, let us denote a column vector of a given data matrix (X_{kj}), k=1,...,N; j=1,...,n by θ_j and the complement of θ_j by

$$\bar{\theta}_j = 1 - \theta_j$$

Then the proportions of rights and wrongs, respectively, for item j are expressed by

$$P(\theta_j) = (1/N)\sum_{k=1}^{N} x_{kj} \quad ,$$

and

$$P(\bar{\theta}_j) = (1/N)\sum_{k=1}^{N}(1 - x_{kj})$$

$$= 1 - P(\theta_j).$$

The proportion $P(\theta_i, \theta_j)$ of subjects getting both items i and j right is given by

$$P(\theta_i, \theta_j) = (1/N)\sum_{k=1}^{N} x_{ki}x_{kj} = n_{11}/N$$

The proportion of subjects getting item i wrong and item j right is

$$P(\bar{\theta}_i\ \theta_j) = (1/N)\sum_{k=1}^{N}(1 - x_{ki})x_{kj} = n_{01}/N \quad .$$

Similarly,

$$P(\theta_i, \bar{\theta}_j) = (1/N) \sum_{k=1}^{N} x_{ki}(1 - x_{kj}) = n_{10}/N,$$

$$P(\bar{\theta}_i, \bar{\theta}_j) = (1/N) \sum_{k=1}^{N} (1 - x_{ki})(1 - x_{kj}) = n_{00}/N$$

These proportions are summarized in the following contingency table of the items i and j.

Insert Figure 2 about here

The variance, covariance and correlation of the four vectors, $\theta_i, \theta_j, \bar{\theta}_i, \bar{\theta}_j$ are given below:

* $\sigma^2(\theta_j) = P(\theta_j)\, P(\bar{\theta}_j) = P(\theta_j)[1 - P(\theta_j)]$
* $\sigma(\theta_i, \theta_j) = P(\theta_i, \theta_j) - P(\theta_i)\, P(\theta_j)$
* $\rho(\theta_i, \theta_j) = [P(\theta_i, \theta_j) - P(\theta_i)P(\theta_j)]/\sqrt{P(\theta_i)P(\bar{\theta}_i)P(\theta_j)P(\bar{\theta}_j)}$

Symbolizing item i as O_i (for "observable No. i"), the definition of Takeya's relation ---> between O_i and O_j is that

$O_i \dashrightarrow O_j$ iff $P(\bar{\theta}_i, \theta_j) < \mu P(\bar{\theta}_i)\, P(\theta_j)$

where μ is a constant between 0 and 1, and usually $0.4 < \mu < 0.6$. His relation is based on a criterion of "negative dependence" between vectors $\bar{\theta}_i$ and θ_j. If the proportion in cell (0,1) -- O_i wrong and O_j right -- is less than about one-half of what would be expected when responses to items i and j are independent, then $O_i \dashrightarrow O_j$. He further defines the following:

D1: $O_i \dashrightarrow O_j$ and $O_j \dashrightarrow O_j \Leftrightarrow O_i \sim O_j$
(items i and j are "equivalent")

D2: $O_i \dashrightarrow O_j$ and $O_j \not\dashrightarrow O_j \Leftrightarrow O_i \rightarrow O_j$
(item i is "easier" than item j)

D3: $O_i \not\dashrightarrow O_j$ and $O_j \dashrightarrow O_i \Leftrightarrow O_j \rightarrow O_i$
(item j is "easier" than item i)

D4: $O_i \not\dashrightarrow O_j$ and $O_j \not\dashrightarrow O_j \Leftrightarrow O_i \perp O_j$
(item i is orthogonal to item j)

For the sake of mathematical convenience, he denoted the complement of $P(\bar{\theta}_i, \theta_j)/P(\bar{\theta}_i)\, P(\theta_j)$ by r^*_{ij} and named it the <u>Coefficient of Ordinality</u>:

i \ j	1	0	total
1	$P(\theta_i,\theta_j)$	$P(\theta_i,\overline{\theta}_j)$	$P(\theta_i)$
0	$P(\overline{\theta}_i,\theta_j)$	$P(\overline{\theta}_i,\overline{\theta}_j)$	$P(\overline{\theta}_i)$
total	$P(\theta_j)$	$P(\overline{\theta}_j)$	1

Figure 2: The Contingency Table of Items i and j

11

$$r^*_{ij} = 1 - P(\bar{\theta}_i, \theta_j)/P(\bar{\theta}_i)\, P(\theta_j)$$

The Coefficient of Ordinality r^*_{ij} can be rewritten as the ratio of ϕ coefficient and the quantity

$$[P(\bar{\theta}_i)\, P(\theta_i)/P(\theta_i)\, P(\bar{\theta}_j)]^{1/2}$$

(1) $$r^*_{ij} = \rho(\theta_i, \theta_j) \Big/ \sqrt{P(\bar{\theta}_i)\, P(\theta_j)/P(\theta_i)\cdot P(\bar{\theta}_j)}$$

Coefficient of Ordinality r^*_{ij} enables us to see the difficulties of items i and j as well as how the ϕ coefficient relates interactively to the order of the two items. Krus & Bart (1973) and Bart & Airasian (1975) defined the dominance relation of two items by taking the probability, $P(\bar{\theta}_i, \theta_j)$ into account. Item i is said to be dominated by item j if $P(\bar{\theta}_i, \theta_j) \leq \varepsilon$ for some small number ε, say $0.02 \leq \varepsilon \leq 0.04$. Bart, et al.'s dominance relation does not reflect either $\rho(\theta_i, \theta_j)$ or $\rho(\bar{\theta}_i, \bar{\theta}_j)$. Thus, the dominance relations of items having a high ϕ value with similar difficulty levels often cannot be defined. Also, two items that are independent can have a consistent dominance relation if their difficulty levels are clearly different.

Takeya's definition of r^*_{ij} enables us to avoid these conflicts. He proved a series of properties regarding the relations among item difficulty, $\rho(\theta_i, \theta_j)$ and dominance of items. They are listed below without proofs.

Property 1: If $O_i \rightarrow O_j$ (by D2) then $\rho(\theta_i, \theta_j) > 0$ and $P(\theta_i) > P(\theta_j)$.

Property 2: If $\rho(\theta_i\, \theta_j) > 0$ and $P(\theta_i) > P(\theta_j)$ then $r^*_{ij} > r^*_{ji}$

Property 3: A set of items whose elements satisfy the circularity $O_{i_1} \rightarrow O_{i_2}, O_{i_2} \rightarrow O_{i_3}, \ldots, O_{i_m} \rightarrow O_{i_1}$ does not exist.

Property 4: If $P(\theta_i) \leq P(\theta_j)$ and $O_i \longrightarrow O_j$ then $O_i \sim O_j$ (D4).

Property 5: If $O_i \sim O_j$, then $\rho(\theta_i, \theta_j) > 1 - \mu$.

A matrix called the Item Relation Structure Analysis Matrix (IRSA Matrix) is formed by calculating r^*_{ij} for all pairs of i and j and if r^*_{ij} is larger than a constant, say .60, replacing the (i,j)-cell by 1, otherwise by 0.

Examination of Item Structure of Fraction Test

Item Relation Structure Analysis Matrix: IRSA Matrix

We have constructed tests of 48 fraction addition and 42 subtraction problems which are expected to be capable of diagnosing erroneous rules resulting from incomplete knowledge or misconceptions occurring at one or more components in the procedural tree given in Figure 3. A detailed description is given in Research Report 81-6 (Klein, et al., 1981). The two tests consist of two parallel subtests in which one of each pair of purportedly parallel items is placed in each subtest. It was noted that constructing two parallel items in terms of having identical procedural steps (number of reducing needed to get the right answer, or obtaining the least common denominator of the two fractions by a prime-factor approach) requires great care and attention. A couple of hundred bugs discovered in our previous studies of signed-number arithmetic (Tatsuoka, et al., 1980; Tatsuoka & Tatsuoka, 1981; Birenbaum & Tatsuoka, 1980, 1981) have shown the necessity of special attention to the thorough examination of detailed procedural steps.

Insert Figure 3 about here

The conventional item analysis using the variances and correlational relations of items and the total scores will not work for tests aimed at diagnostic use. Examination of item performances with respect to the expected operational functioning of each item when they were constructed must be carried out into the level of individualized behavior of each item and the interactions among items. IRSA method seems to provide the information of item behaviors with the help of a digraph representation of the dominance relations among items. Moreover, Takeya's r^*_{ij} has algebraic relations with traditional statistics such as reliability, Cliff's consistency index, Loevengier's homogeneity index (Takeya, 1981).

We adopted IRSA to investigate the fraction test items. Table 1 is the Item Relation Structure Matrix in which the 24-items from the first subtest of the 48-item addition test were sorted by their p-values (proportion correct).

Insert Table 1 about here

Table 2 provides the p-values of the 24 items.

Insert Table 2 about here

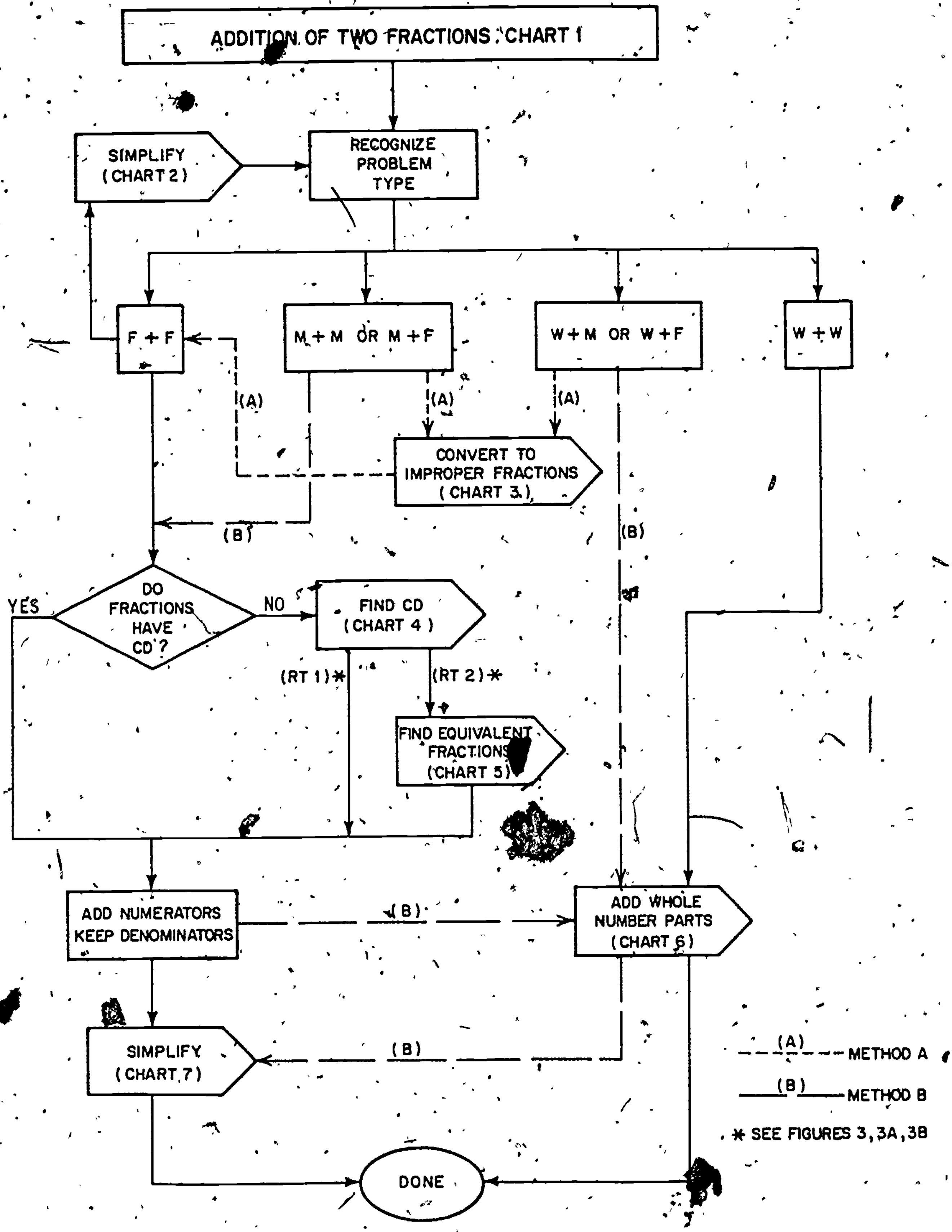

FIGURE 3: A Procedural Network for Adding Two Fractions

Table 1

Item Relation Structure Analysis, (IRSA) Matrix of the
First 24 Fraction Addition Problems

Item	Item	2	5	12	4	9	6	11	14	19	3	41	31	82	22	0	71	22	12	3	81	51	61	71	0
1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	24	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
5	9	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	6	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	11	1	1	0	1	1	0	1	0	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	0
8	14	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	19	1	1	0	0	1	1	1	0	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0
10	3	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0
11	4	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	0	0	0	1	1	1	0	0
12	13	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	0	0
13	18	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0
14	22	1	1	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
15	20	1	1	0	1	0	1	1	0	0	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0
16	7	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	0	0	1	1	0	0	0
17	12	1	1	0	0	0	1	1	0	1	1	1	1	1	0	0	1	1	0	0	1	0	0	0	0
18	21	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
19	23	1	1	0	1	0	1	1	0	1	1	1	0	1	1	1	0	0	0	1	1	0	0	0	0
20	8	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0
21	15	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0
22	16	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0
23	17	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1

Table 2

Difficulties, $P(\theta_i)$ of the First 24

Fraction Addition Problems

N = 154

	Items	$P(\theta_i)$
1	2	.142857
2	5	.240260
3	1	.246753
4	24	.305195
5	9	.324675
6	6	.337662
7	11	.357143
8	14	.357143
9	19	.370130
10	3	.376623
11	4	.376623
12	13	.376623
13	18	.376623
14	22	.376623
15	20	.383117
16	7	.389610
17	12	.389610
18	21	.389610
19	23	.389610
20	8	.402597
21	15	.409091
22	16	.454545
23	17	.571429
24	10	.584416

The scoring of the test was obtained by matching each part of an answer to the corresponding part of the right answer. By "parts" is here meant the whole number part (W) of the answer, and the denominator and numerator of the fractional part (F). For example, if the right answer requires reducing the fraction to its lowest terms, and a student did not carry out the reducing but got the whole number part right, then a three-element binary vector (1 0 0) is stored as the student's score in the PLATO system. The item difficulties shown in Table 2 are based on the customary right/wrong score multiplied by the three elements of the score vector such as (1 0 0).

In order to construct a digraphed tree from the IRSA matrix, we must first extract useful, systematic information existing among various subsets of items. As it stands, the IRSA matrix in Table 1 is too complicated to extract chains because there are too many items, and the dimensionality (in the Guttman-scale sense) of the dataset is large -- even though, in principle, it may be used like an adjacency matrix, with one exception as noted below.

Super-Order of Items

Takeya's item relation "—>" does not satisfy the transitivity law which is crucial for extracting a chain from the original set of items. Although he extended his theory to the case of three or more items and defined the coefficient of ordinality for a finite number of items, transitivity still does not necessarily hold unless inequality (2) is satisfied. That is,

$$O_i \longrightarrow O_j \text{ and } O_j \longrightarrow O_m \text{ imply } O_i \longrightarrow O_m \text{ provided}$$

$$\rho(\theta_i, \theta_m) > \rho(\theta_i, \theta_j)\rho(\theta_j, \theta_m)/(1-\mu) \tag{2}$$

It can be seen in real data that inequality (2) sometimes does not hold. To cope with this problem we proceed as follows. [Takeya is not explicit about the actual algorithm he uses to extract his chains, which is presumably a trade secret of his company. What we describe below may therefore not be the most efficient algorithm, but it does work.]

For each item O_i, we define its antecedent set A_i as the set of all items O_j that have the relation O_j —> O_i. (The members of A_i are those items O_j in the IRSA matrix that have unit entries in the column representing O_i.) Next, for each item O_i, in the order in which they occur from top to bottom in the IRSA matrix, we list the item numbers m, of those items O_m that occur to the right of O_i and whose antecedent set

Insert Table 3 about here

A_m is a subset of A_i. Table 3 shows such a listing for the present example. Next, we reconstrue this listing to give new, formal definitions of A_i as sets of the integers standing in the line headed by the subscript i of A_i. Thus, for example, the fourth row of Table 3, headed by 24, is construed as the formal redefinition $A_{24} = \{15, 16, 17, 10\}$. [We realize that the numerals in each row themselves originally stood for antecedent sets with those numerals as subscripts. It may therefore seem that we are talking about "sets of sets that are subsets of the former" in a manner reminiscent of Russell's paradox! However, we are merely making formal redefinitions of each A_i for the sole purpose of facilitating the description of the next and final step. We are not asserting, for example, that $A_{24} = \{A_{15}, A_{16}, A_{17}, A_{10}\}$ and committing the fallacy of confusing two levels of "set-hood". Nor are we making any conflicting redefinition like $A_{24} = \{O_{15}, O_{16}, O_{17}, O_{10}\}$, which contradicts the original (and true) definition of A_{24}. Finally, starting from the bottom of Table 3 we trace sequences, $A_{i_1}, A_{i_2}, \ldots, A_{i_s}$, of antecedent sets that successively subsume the ones to the left; i.e., $A_{i_1} A_{i_2} \subset \cdots \subset A_{i_s}$. In the present example, the 16 sequences shown below are found. (Note that five of the lines have two-character headings like c,c´. Each of these lines lists two sequences differing only in whether the last set is A_2 or A_5, as do the sequences in lines a and a´.)

a, a´ $A_{10} \subset A_{17} \subset A_{16} \subset A_{15} \subset A_8 \subset A_{18} \subset A_3 \subset A_{11} \subset A_2$ (or A_5)

b $A_{10} \subset A_{17} \subset A_{16} \subset A_{15} \subset A_{13} \subset A_4 \subset A_2$

c, c´ $A_{10} \subset A_{17} \subset A_{16} \subset A_{15} \subset A_{20} \subset A_{11} \subset A_2$ (or A_5)

d, d´ $A_{10} \subset A_{17} \subset A_{12} \subset A_7 \subset A_{18} \subset A_3 \subset A_{11} \subset A_2$ (or A_5)

e, e´ $A_{10} \subset A_{17} \subset A_{19} \subset A_{11} \subset A_2$ (or A_5)

Table 3

Antecedent Sets A_i Construed as Sets of Integers m

Such That $A_i \subset A_m$

i	"Elements" of A_i
2	24 6 11 14 19 3 4 13 18 22 20 7 12 21 23 8 15 16 17 10
5	9 6 11 14 19 3 18 20 7 12 23 8 15 16 17 10
1	14 17 10
24	15 16 17 10
9	17 10
6	17 10
11	19 3 18 20 7 12 23 8 15 16 17 10
14	17 10
19	17 10
3	18 7 12 8 15 16 17 10
4	13 12 8 15 16 17 10
13	12 15 16 17 10
18	7 12 8 15 16 17 10
22	21
20	15 16 17 10
7	12 16 17 10
12	17 10
23	17 10
8	15 16 17 10
15	16 17 10
16	17 10
17	10
10	10

f, f′ $A_{10} \subset A_{17} \subset A_6 \subset A_2$ (or A_5)

g $A_{10} \subset A_{17} \subset A_9 \subset A_5$

h $A_{10} \subset A_{17} \subset A_{14} \subset A_1$

o, o′ $A_{10} \subset A_{17} \subset A_{23} \subset A_{11} \subset A_2$ (or A_5)

r $A_{21} \subset A_{22} \subset A_2$

The subscripts of the A′s in each of these sequences identify the items that constitute a chain of approximately unidimensional items. What we have accomplished by the set of operations described above is to identify subsets of items for which transitivity of the relation $\rightarrow$ holds.

Appendix I presents the 48-item test administered to 154 local junior-high students. Item No′s. 21 and 22 in the first subtest (the first 24 items) involve the addition of a mixed number and a whole number; these items are in chain r. It is obvious that W + F, W + M types usually don′t involve the operation of finding the least common denominator, so they form a chain independent from other types of skills unless a student uses Method A (converts to a F + F type). M + M type items (13, 4, 12, 6, 9, 1) are not included in the longest chains, a and a′, in which the types of items are mostly F + F. Simplifying a fraction to a mixed number, converting a mixed number to a fraction seem to cause children a considerable amount of trouble. Items 2 and 5 involve relatively prime numbers in their denominators as well as the simplification of the final answer. The shaded nodes in the digraph of Figure 4 stand for the items whose final answer needs simplification. It is fairly clear that if a student has some misconception about the simplication procedure, then he/she tends to repeat the same mistake until it is corrected. Thus, the items requiring simplification tend to be located toward the end of the chains in Figure 4. The items that are

Insert Figure 4 about here

similar types such as W + M or W + F -- adding a whole number and a mixed (or fraction) — tend to cluster together, but 21 and 22 which are of the M + W type don′t have any relation with other types of items.

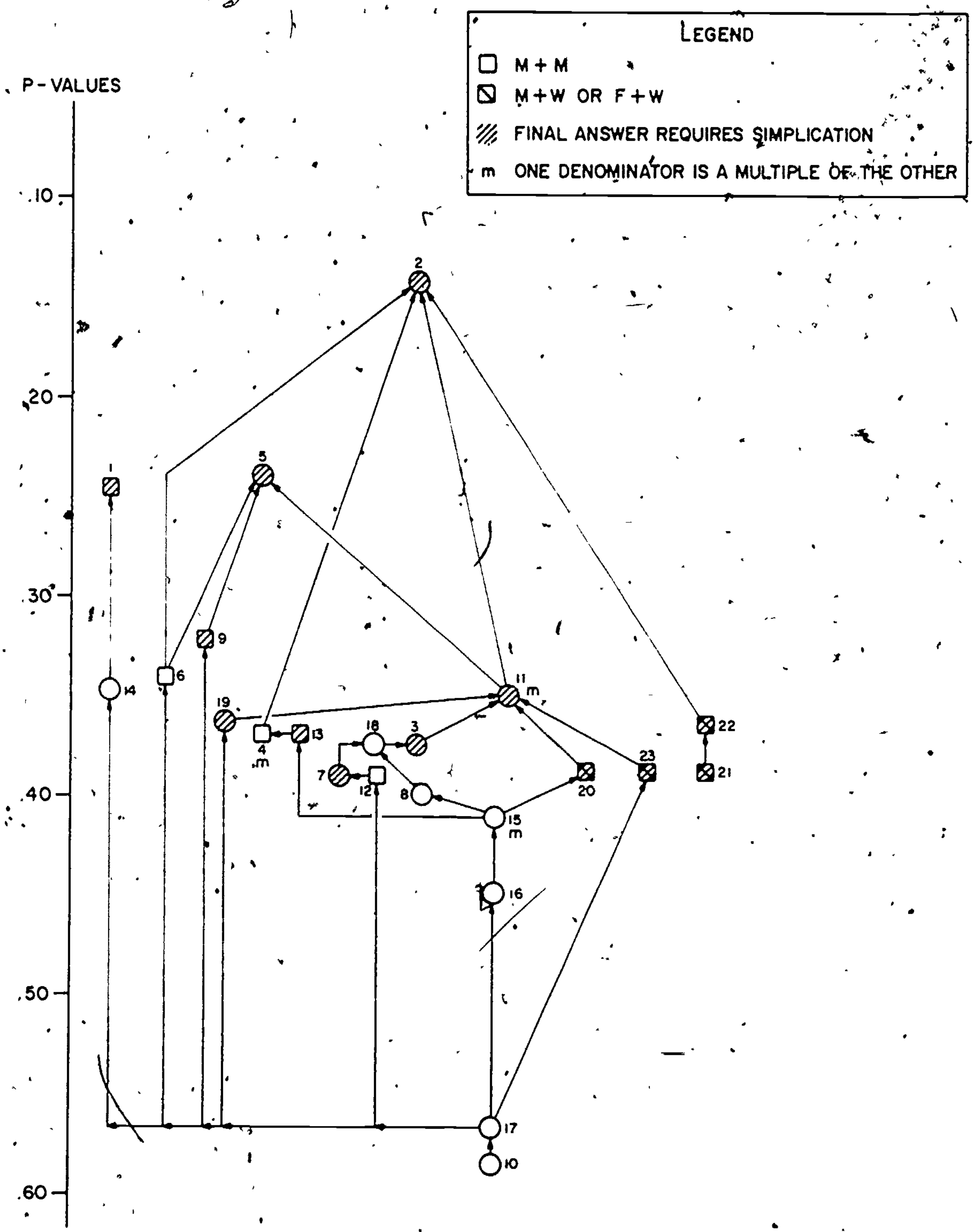

FIGURE 4: A Directed Graph of the First 23 Addition Items

Items 20 and 23 behave differently from 21 and 22. More extensive error analyses should provide a resolution of such apparent anomalies.

Summary and Discussion

We have described innovative techniques that will help in constructing error-diagnostic tests and examining the appropriateness and necessity of each item in order to enable the test to provide a specific description of misconceptions. Brown & Burton's (1978) artificial intelligence approach (BUGGY) is innovative and very impressive. However, as the number of discovered bugs increases enormously, the algorithm of a computer program becomes more and more complicated. Apart from their approach, the authors (Tatsuoka, et al., 1980; Tatsuoka & Tatsuoka, 1980; Tatsuoka & Linn, 1981) have been working to develop psychometric models that should have capabilities comparable to BUGGY and, moreover, will be able to handle several hundred bugs with a probabilistic approach.

A technique to investigate the item structure with respect to the roles of each item in determining the student's misconceptions will be in great demand for error-diagnostic testing. Item Relation Structured Analysis was tried out to obtain our much needed information, and a new chain-formation procedure was introduced in this paper. The result more or less confirmed that our item-construction procedure was on the right track, but further and more extensive investigations from different angles -- such as the examination of whether two purportedly parallel items are really parallel -- will be needed.

References

Airasian, P. W., & Bart, W. M. Ordering Theory: A new and useful measurement model. Journal of Educational Technology, May 1973, 56-60.

Bart, W. M., & Krus, D. J. An ordering-theoretic method to determine hierarchies among items. Educational and Psychological Measurement, 1973, 33, 291-300.

Birenbaum, M., & Tatsuoka, K. K. The use of information from wrong responses in measuring students' achievement (Research Report 80-1). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1980.

Birenbaum, M., & Tatsuoka, K. K. Effect of different instructional methods on error tyes and the underlying dimensionality of the test, Part I (Research Report 81-3). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, August 1980.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.

Cliff, N. A theory of consistency or ordering generalizable to tailored testing. Psychometrika, 1977, 42, 375-399.

Guttman, L. L. Studies in social psychology in World War II. In S. A. Stouffer (Ed.), Measurement and Prediction, Vol. IV. Princeton: Princeton University Press, 1950.

Krus, D. J. Order analysis of binary data matrices. Los Angeles: Theta Press, 1975.

Klein, M. F., Birenbaum, M., Standiford, S. N., & Tatsuoka, K. K. Logical error analysis and construction of tests to diagnose student "bugs" with addition and subtraction of fractions (Research Report 81-6). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, November 1981.

Loevinger, J. The technic of homogeneous tests compared with some aspects of "scale analysis" and "factor analysis." Psychological Bulletin, 1948, 45, 507-529.

Mokken, R. J. A theory and procedure of scale analysis: With applications in political research. The Hague: Mouton, 1971.

Sato, T. The construction and interpretation of S-P tables. Tokyo: Meiji Tosho, 1975 (in Japanese).

Takeya, M. A study on item relational structure analysis of criterion referenced tests. Unpublished Ph.D. dissertation. Tokyo, Waseda University, June 1981.

Tatsuoka, K. K., & Birenbaum, M. The danger of relying solely on diagnostic adaptive testing when prior and subsequent instructional methods are different, CERL Report E-5. Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, March 1979.

Tatsuoka, K. K., & Birenbaum, M. The effect of different instructional methods on achievement tests. Journal of Computer-Based Instruction, in press.

Tatsuoka, K. K., Birenbaum, M., Tatsuoka, M. M., & Baillie, R. A psychometric approach to error analysis on response patterns (Research Report 80-3). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1980.

Tatsuoka, K. K., & Linn, R. L. Indices for detecting unusual response patterns: Links between two general approaches and potential applications (Research Report 81-5). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, August 1981.

Tatsuoka, K. K., & Tatsuoka, M. M. Detection of aberrant response patterns and their effect on dimensionality (Research Report 80-4). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1980.

Tatsuoka, K. K., & Tatsuoka, M. M. Spotting erroneous rules of operation by the Individual Consistency Index (Research Report 81-4). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1981.

Wise, S. L. A modified order-analysis procedure for determining unidimensional items sets. Unpublished doctoral dissertation, University of Illinois, 1981.

Appendix I

The First 24 Items in a 48-Item Addition Problems in Fractions Test

NAME ____________________

GRADE ____________________

TEACHER ____________________

SHOW ALL YOUR WORK SO THAT YOU CAN RECIEVE PARTIAL CREDIT.

1. $2\frac{8}{6} + 3\frac{10}{6} =$

2. $\frac{2}{5} + \frac{12}{8} =$

3. $\frac{8}{5} + \frac{6}{5} =$

4. $2\frac{1}{2} + 4\frac{2}{4} =$

5. $\frac{1}{2} + 1\frac{10}{7} =$

6. $3\frac{5}{7} + 4\frac{6}{7} =$

7. $\frac{3}{5} + \frac{7}{5} =$

8. $\frac{1}{3} + \frac{1}{2} =$

9. $1\frac{4}{7} + 1\frac{12}{7} =$

10. $\frac{3}{5} + \frac{1}{5} =$

11. $\frac{3}{4} + \frac{1}{2} =$

12. $2\frac{5}{9} + 1\frac{1}{9} =$

13. $3\frac{1}{6} + 2\frac{3}{4} =$

14. $\frac{15}{35} + \frac{10}{35} =$

15. $\frac{1}{2} + \frac{3}{8} =$

16. $1\frac{2}{5} + \frac{3}{5} =$

17. $\frac{1}{4} + \frac{3}{4} =$

18. $\frac{4}{15} + \frac{1}{10} =$

19. $\frac{4}{5} + \frac{3}{5} =$

20. $3 + \frac{10}{5} =$

21. $1\frac{3}{9} + 2 =$

22. $3 + 1\frac{5}{2} =$

23. $\frac{12}{8} + 1$

24. $2\frac{1}{3} + 3\frac{1}{4} + 2\frac{2}{6} =$