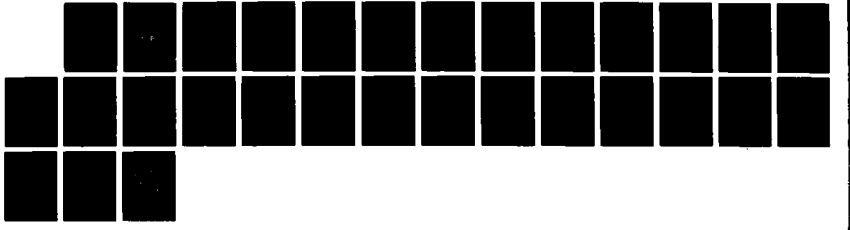
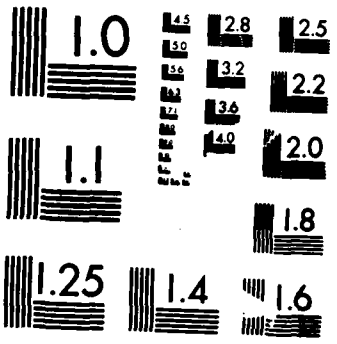
AD-A183 173 ITEM RESPONSE THEORY LATENT CLASSES AND RULE SPACE(U) 1/1
ILLINOIS UNIV AT URBANA COMPUTER-BASED EDUCATION
RESEARCH LAB K K TATSUOKA ET AL. 15 JUL 87 87-2-ONR
UNCLASSIFIED N00014-82-K-0604 F/G 12/9 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

(12)

AD-A183 173

Item Response Theory, Latent Classes and Rule Space

KIKUMI TATSUOKA

MAURICE TATSUOKA

UNIVERSITY OF ILLINOIS

DTIC
ELECTE
JUL 3 0 1987
S D
CD

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution unlimited

87 7 27 040

unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ONR 87-2	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Computer-based Ed Res Lab University of Illinois 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Personnel and Training Research Programs
6c. ADDRESS (City, State, and ZIP Code) 252 ERL 103 S. Mathews Ave. Urbana, IL 61801	7b. ADDRESS (City, State, and ZIP Code) Office of Naval Research (Code 1142PT) 800 N. Quincy St. Arlington, VA 22217-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-82-K-0604
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61153N; PROJECT NO. RR04204; TASK NO. RR04204-01; WORK UNIT ACCESSION NO. NR150-495

11. TITLE (Include Security Classification)
Item Response Theory, Latent Classes and Rule Space: Final Report

12. PERSONAL AUTHOR(S) Kikumi K. Tatsuoka & Maurice M. Tatsuoka

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 1982 TO 1986	July 15, 1987	26

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Item Response Theory, Latent Classes, Rule Space, Cognitive Error Diagnoses, Unidimensionality
05	09		

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

This study demonstrated that item response theory, latent class models, and the rule space model introduced by Tatsuoka (1985) and Tatsuoka and Tatsuoka (1987) are algebraically related. Specifically, it was shown (1) that IRT functions may actually be regarded as the conditional density functions of item scores for a special latent class representing the "null state" of knowledge (i.e., the state that would "ideally" produce a response vector of all zeros); and (2) that estimates of the item parameters of IRT functions can be determined from the union of several latent classes with the following property: when their response vectors are mapped into rule space, the centroids of these projections lie approximately along the first principal axis of the union set.

Bug distributions, which are density function of the numbers of slips away from
(contd. on reverse)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Charles E. Davis	22b. TELEPHONE (Include Area Code) (202) 696-4046
	22c. OFFICE SYMBOL ONR 1142PT

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. Abstract (contd.)

the ideal rule-generated response patterns, play an important role in interrelating IRT and latent-class models; they in fact hold the key to the development of a general theory of rule space that includes these two models as special cases. Futhermore, bug distributions form the basis for developing new indices that measure the stability of states or rules and the consistency with which a particular rule applied with no intrusion of slips.

ABSTRACT

This study demonstrated that item response theory, latent class models, and the rule space model introduced by Tatsuoka (1985) and Tatsuoka and Tatsuoka (1987) are algebraically related. Specifically, it was shown (1) that IRT functions may actually be regarded as the conditional density functions of item scores for a special latent class representing the "null state" of knowledge (i.e., the state that would "ideally" produce a response vector of all zeros); and (2) that estimates of the item parameters of IRT functions can be determined from the union of several latent classes with the following property: when their response vectors are mappped into rule space, the centroids of these projections lie approximately along the first principal axis of the union set.

Bug distributions, which are density function of the numbers of slips away from the ideal rule-generated response patterns, play an important role in interrelating IRT and latent-class models; they in fact hold the key to the development of a general theory of rule space that includes these two models as special cases. Furthermore, bug distributions form the basis for developing new indices that measure the stability of states or rules and the consistency with which a particular rule is applied with no intrusion of slips.

Accesion For
NTIS CRA&I
DTIC TAB
Unannounced
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

QUALITY INSPECTED 4

Recent advances in cognitive theory provide new insights into human thinking and learning processes. Linn (1985) pointed out that it is important to develop a new theory and measurement technique in order to measure volatile learning activities described in Glaser (1985) and assess cognitive skill acquisition. Glaser (1985) summarized the main objectives of assessing new achievement measures into four categories: 1) Diagnosing the principles of performance; 2) Assessing the theory changes; 3) Evaluating a structure or representation of problems; and 4) Assessing the automaticity of performance skills. A modern measurement theory must be developed by taking the four objectives into account. This implies that it is necessary to establish a concept of item construction that is different from the classical foundation. Traditionally, item construction originated from the evaluation of content validity--how a test covers subject matter and situations. Again, Glaser (1985) suggested that test items could be comprised of two elements--"information that needs to be known and information about the conditions under which use of this knowledge is appropriate." As for the former element, there are various stages of competence in students knowledge, including cognitive skills. Also, it is important to assess what knowledge structure the students have. Greeno (1980) pointed out that the acquisition of declarative and procedural knowledge is usually an objective of instruction, but that strategic knowledge that enables one to set goals and subgoals and to form plans for attaining goals is not explicitly taught. Different item types often require the students to decide which solution path should be taken, and what should be done first to reach the final answer.

Many erroneous rules discovered in past research (Brown & Burton, 1978; VanLehn, 1983; Tatsuoka & Tatsuoka, 1981) indicated many erroneous rules originated from a lack of the strategic skill described by Greeno. Therefore, new measurements must include the information for prescribed diagnosis of students' erroneous rules or sources of misconception. The new test design must be capable of reflecting and discriminating between the different knowledge structures possessed by individuals. Each structure requires its unique strategies to set subgoals and goals and to find solution paths. As a result, different structures produce different sets of erroneous rules; some rules may be included in both the structures but the others are included in just one of them. The modern measurement theory must be able to discriminate one knowledge structure from another. The third condition "theory change" is stated as "hypotheses testing." When learning takes place, students test their hypotheses and then evaluate, examine, and modify current theories on the basis of new information. It is not unusual that many students change their erroneous rules one to another before reaching the mastery stage. Measurements of new kinds of tests must capture the traces of these performance changes in detail in order to increase educational utility of responses to the test. The goals to be attained in modern measurement theory are not easy. Apparently, the technical barriers, as Linn states, are high and the traditional theories of educational measurement and testing have only limited power, or are simply inapplicable to the new measures.

In this paper, the pros and cons of two representative test models, Item Response Theory (Lord & Novick, 1968) and Latent Class (Lazarsfeld

& Henry, 1968) will be discussed with respect to the demands of modern measurement theory and their interrelationships with rule space will be discussed. Discussion will be focussed on their modeling assumptions, and conditional density functions of latent rules (classes or groups). It will also be shown that IRT becomes a special case of latent rule (classes or groups).

A cluster of response patterns around Rule R

If a student applies his/her erroneous rule with perfect consistency to the items in the test, then his/her responses to the test will be perfectly matched with the responses generated by a computer program. We call such systematic errors erroneous rules or rules. A correct rule will, by definition, produce the right answer to all the items. Although wrong rules sometimes may produce the right answer to some subset of the test items, it is very unlikely that they will produce the right answer to all the items. We further assume that the test items are carefully constructed so that the important, predicted common errors can be expressed by unique item response patterns of ones and zeros. Therefore, rule R can be represented by a binary vector $\underset{\sim}{R} = (r_1, r_2, \ldots r_n)$. However, actual students performances on the test items are unlikely to be perfectly consistent and are subject to random errors or slips due to carelessness or uncertainty that always affect the outcomes of performances on a test. Even if a student possesses some systematic error, it is rare to have the response pattern perfectly matched with the pattern theoretically generated by its algorithm (VanLehn, 1983; Tatsuoka, 1984). Some systematic errors may have a

tendency to produce more slips while other rules produce fewer slips. Some items may be prone to produce more slips than other items. Thus, it would not be realistic to assume that all the items have equal slip probabilities.

Bug Distributions

Tatsuoka & Tatsuoka (1987) derived the theoretical distribution of observed slips, and called it "bug distribution." First, the probability of having a slip on item j (j=1,2,...,n) is denoted by p_j for item j

(1) $P_{j|R}$ = Prob (having a slip on item $j|R$) = Prob ($u_j = 1|R$)

where u_j is a random variable such that $u_j = 1$ if a slip occurs on item j and $u_j = 0$ if not, and Rule R is a vector $\underset{\sim}{R} = (r_1, r_2, \ldots, r_n)$.

(2)
$$\begin{cases} u_j = 1 & \text{if a slip occurs (i.e., if } x_j \neq r_j) \\ u_j = 0 & \text{otherwise} \qquad\quad (\text{i.e., if } x_j = r_j) \end{cases}$$

More succinctly, u_j may be defined as

(2a) $$u_j = |r_j - x_j|$$

Given the reasonable assumption that slips occur independently across items, the bug distribution of rule R follows a compound binomial distribution with different slip probabilities for the items

(3) $$\text{Prob (having up to s slips}|R) = \sum_{m=0}^{s} \left[\sum_{\Sigma u_j = m} \left\{ \prod_{j=1}^{n} P_j^{u_j} (1-P_j)^{1-u_j} \right\} \right]$$

Since R can be any rule, the number of slips from the correct rule $\underset{\sim}{R} = (1,1,\ldots 1) = \underset{\sim}{1}$ also follows a compound binomial distribution with

slip probabilities $p_{j|\underset{\sim}{1}} = \text{Prob}(u_j = 0_j|\underset{\sim}{1})$, $j = 1,\ldots,n$. If the elements of Rule R are zeros, $\underset{\sim}{0} = (0,0,\ldots 0)$, then the slip probabilities from this wrong rule will be given by $p_{j|\underset{\sim}{0}} = \text{Prob}(u_j = 1|\underset{\sim}{0})$, $j = 1,\ldots,n$.

Property 1. The slip probabilities of the bug distribution are determined by the logistic function of Item Response Theory. The slip probabilities $p_{j|R}$ are given by Equation 4

$$(4) \qquad P_{j|R} = \text{Prob}(u_j = 1;\underset{\sim}{R}) = \text{Prob}(x_j \neq r_j;R) = r_j Q_j(\theta_R) + (1 - r_j)P_j(\theta_R)$$

where x_j is the observed score of item j and $P_j(\theta_R)$ is the IRT function at the θ level associated with rule R.

Suppose Rule R corresponds to a vector $\underset{\sim}{R} = (1111000)$, the first four elements being ones and the others, are zeros. The random variable u_j will be 1 if a slip from r_j occurs and zero if not. If a student's performance on the seven items results in a response pattern of two slips away from $\underset{\sim}{R}$, then two items have different values from the elements of vector $\underset{\sim}{R}$. Suppose the two slips occurred on items 1 and 7, then the corresponding response pattern will be x = (0111001). The middle member of 4 can be rewritten as follows:

$$(5) \qquad \text{If } r_j = 1, \text{ then } \text{Prob}(x_j \neq r_j;R) = \text{Prob}\,(x_j = 0;R)$$
$$\text{If } r_j = 0, \text{ then } \text{Prob}(x_j \neq r_j;R) = \text{Prob}(x_j = 1;R)$$

It is known that the probability of score 1 for item j is the logistic function $p_j(\theta_R)$. $\text{Prob}(x_j = 1;R) = p_j(\theta_R)$, thus Equation 5 can be written

in the single Equation 4, viz.:

$$\text{Prob}\ (x_j \neq r_j;R) = r_j Q_j(\theta_R) + (1 - r_j)P_j(\theta_R).$$

From Equations 2, 3, 4, and 5, the slip probability is given by a weighted mean of r_j and $1_j - r_j$, as follows;

$$P_{j|R} = r_j Q_j(\theta_R) + (1 - r_j)P_j(\theta_R). \tag{6}$$

Alternatively, to simplify the notation, we may write

$$P_{j|R} = |r_j - P_j(\theta_R)| \tag{6a}$$

Equation (6) shows that any slip probability p_j is a function of θ. In order to emphasize this fact, p_j will be denoted by $S_j(\theta)$ hereafter.

The conditional distribution function of number of slips given by Equation 1 will be rewritten in terms of $S_j(\theta)$, in Equation 7 by replacing p_j by $S_j(\theta)$, and Equation 8 is the generating function of expression (7). (We omit the subscript R in θ_R as understood.)

$$\text{Prob (having up to s slips from } \underset{\sim}{R}) = \left[\sum_{m=0}^{s}\left\{\sum_{\Sigma u_j = m}\prod_{j=1}^{n} S_j(\theta)^{u_j}(1 - S_j(\theta))^{1-u_j}\right\}\right. \tag{7}$$

$$g(\theta;R) = \prod_{j=1}^{n}\{S_j(\theta) + (1 - S_j(\theta))\} \tag{8}$$

The expectation and variance of the number of slips from rule R are given by (9) and (10),

$$\mu_R = \sum_{j=1}^{n} S_j(\theta) = \sum_{r_j=0} P_j(\theta) + \sum_{r_j=1} Q_j(\theta) \tag{9}$$

$$\sigma_R^2 = \sum_{j=1}^{n} S_j(\theta)(1 - S_j(\theta)) = \sum_{r_j=0} P_j(\theta)Q_j(\theta) + \sum_{r_j=0} Q_j(\theta)P_j(\theta). \tag{10}$$

A Measure of Rule Stability

The expectation and variance given by equations (9) and (10) are not those of the total score as is customary in conditional density functions of latent rules or classes; rather, they refer to the number of slips. The expectation of the number of slips from a rule is a measure of the instability of rule, since the expectation represents the average number of slips from that rule, and the variance is a measure of the extent to which the number of slips made varies from student to student. For example, for the erroneous rule producing all wrong answers, expressed by the vector $\underset{\sim}{0} = (0,0,\ldots,0)$, the expectation $\sum_{j=1}^{n} P_j(\theta_0)$ will be very small because the values of $P_j(\theta_0)$, $j=1,\ldots,n$ are very small -- nearly zero. Therefore, Rule $\underset{\sim}{0}$ is very stable and slips rarely occur. At the same time, the right rule $\underset{\sim}{1} = (1,1,\ldots,1)$ has the expectation of $\sum_{j=1}^{n} Q_j(\theta_1)$ which also is very small. We can conclude that any students who are in the state of mastery can execute the right rule systematically and the probability of having any slip deviating from the right rule is very small. In general, the mean number of slips from 1 to 0 will be $\sum_{r_j=1} S_j(\theta)$ and the mean number of slips from 0 to 1 will be $\sum_{r_j\neq 0} S_j(\theta)$. The expected number of slips will be $\sum_{j=1}^{n} S_j(\theta)$.

Now let us consider a rule R whose elements are about half ones and half zeros. Then the probability of having slips will be close to .5, (Tatsuoka, 1986), which implies that such rules have a 50% chance of having slips away from its perfect execution. Moreover, since the conditional expectation of the bug distribution is larger around the

mid-range of θ (Tatsuoka, 1986) the number of slips expected to occur there will be fairly large. This means that rules used by many average ability students will tend to have more slips than the rules that very high or very low ability students are likely to use, and the stability of such rules is lower because the probability of having slips is higher for rules espoused by high or low ability students.

A Measure of Rule Consistency

The expectation of the number of slips is given in Equation 9, which can be regarded as a measure of how stable this rule is. The variance given in Equation 10 represents the dispersion or spread of number of slips.

However, consistency of a rule r is a different concept. If a student uses rule r with perfect consistency then the resulting response pattern matches the binary pattern generated by a logically programmed algorithm for rule R. Therefore, the probability of not having any slips can be obtained by setting $u_j = 0$ for $j = 1,\ldots,n$.

$$(11) \qquad \text{Prob (perfect execution of rule R)} = \prod_{j=1}^{n} (1 - S_j(\theta)).$$

The probability obtained from Equation 11 is an index of consistency and represents the probability of systematic execution of rule R. However, the value of the consistency defined in this manner will be extremely small as the number of test items becomes large. The consistency measure must be independent of the test length. The most plausible candidate for the consistency index is

$$(12) \qquad C_R = \left[\prod_{j=1}^{n} (1 - S_j(\theta))\right]^{1/n}$$

which is the geometric mean of the probabilities of not having slips on items 1,2,...,n -- i.e., of the n factors of the right-hand member of Equation 11.

Relationship Between Bug Distribution and Item Response Theory Model

Bug distribution was formulated by taking the notion of slips and slip probabilities. It was assumed that the occurrence of slips was independent across the items and the probability of a slip occurring for each item was denoted by $p_{j|R}$, j=1,...,n. Each item j has its unique chance of having a slip away from r_j and different values of slip probabilities are assumed across the items. Then, the probability of having some finite number of slips away from Rule R was given by Equation 3 with the slip variable u_j defined by Equation 2 or 2a. Derivation of the compound binomial distribution (3) is applicable to any rules--where a set of response patterns resulting from inconsistent application of rule R was introduced as a cluster around rule R in Tatsuoka & Tatsuoka (1987) and was denoted by $\{\underset{\sim}{R}\}$. When an erroneous rule produces wrong answers for all the items in a test, it corresponds to the null vector, $\underset{\sim}{0} = (0,0,\ldots,0)$. In this case the random variable u_j is the same thing as the random variable of item score x_j, so $\text{Prob}(u_j = 1|\underset{\sim}{0}) = \text{Prob}(x_j = 1|\theta_0)$ where θ_0 is the latent ability level of rule $\underset{\sim}{0}$, and the slip probabilities of n items become the logistic functions $P_j(\theta)$ of IRT models. Therefore, it can be said that the IRT model is equivalent to the latent class model associated with the rule that produces the null vector, $\underset{\sim}{0}$. The likelihood functions of rule-$\underset{\sim}{0}$-latent class and IRT are as follows:

$$(13) \qquad L(\underset{\sim}{O}) = \prod_{i\in\{\underset{\sim}{O}\}} \prod_{j} \left[P_j(\theta)^{u_j} (1 - P_j(\theta))^{1-u_j} \right]$$

$$(12) \qquad L(IRT) = \prod_{i=1}^{N} \prod_{j=1}^{n} \left[P_j(\theta)^{x_j} (1 - P_j(\theta))^{1-x_j} \right]$$

where N is <u>not</u> the number of subjects belonging to the cluster of latent rule $\underset{\sim}{O}$, {O}, but is simply the sample size.

A sample whose response patterns are well described by the IRT model (i.e. an "IRT sample") may contain clusters around many rules, including the cluster around the correct rule $\underset{\sim}{1}$. Let R be one of many rules contained in the IRT sample. Then, the likelihood of the bug distribution associated with Rule R is

$$(15) \qquad L(\underset{\sim}{R}) = \prod_{i\in\{R\}} \prod_{j=1}^{n} \left\{ S_j(\theta)^{u_j} (1 - S_j(\theta))^{1-u_j} \right\}.$$

The relationship between the two variables u and x was already given in Equation 2a, viz., $u_j = |r_j - x_j|$. Also, the slip probability of $S_j(\theta)$ is given by Equation 6a with $P_{j|R}$ now rewritten as $s_j(\theta)$. Substituting u_j and $S_j(\theta)$ from these two relations, Equation 16 is obtained,

$$(16) \qquad L(\underset{\sim}{R}) = \prod_{i\in\{\underset{\sim}{R}\}} \prod_{j=1}^{n} |r_j - P_j(\theta_i)|^{|r_j - x_j|} (1 - |r_j - P_j(\theta_i)|)^{(1 - |r_j - x_j|)}$$

Separating the multiplication over j into those factors for which $r_j = 1$ and those for which $r_j = 0$, we get

$$L(\underset{\sim}{R}) = \prod_{i\in\{\underset{\sim}{R}\}} \left\{ \prod_{j\supset r_j=1} [1 - P_j(\theta_i)]^{1 - x_{ji}} [1 - (1 - P_j(\theta_i))]^{1 - (1 - x_{ji})} \right.$$

$$\left. \times \prod_{j\supset r_j=0} |- P_j(\theta_i)|^{1 - x_{ji}} [1 - (1 - |- P_j(\theta_i)|^{1 - |- x_{ji}|} \right.$$

or

$$(17) \qquad L(\underset{\sim}{R}) = \prod_{i\in\{\underset{\sim}{R}\}} \left\{ \prod_{r_j=1} \left[Q_j(\theta)^{1-x_{ji}} P_j(\theta)^{x_{ji}} \right] \left[\prod_{r_j=0} P_j(\theta)^{x_{ji}} Q_j(\theta)^{1-x_{ji}} \right] \right\}.$$

Therefore expression (17) becomes the same as the conventional likelihood expressed in terms of the score variable x_j and IRT function $P_j(\theta)$ upon combining the $r_j = 1$ and $r_j = 0$ cases. Thus, Equation 18 is obtained;

$$(18) \qquad L(\underset{\sim}{R}) = \prod_{i\in\{R\}} \prod_{j=1}^{n} \left\{ P_j(\theta)^{x_{ji}} Q_j(\theta)^{1-x_{ji}} \right\}.$$

Equation 18 is exactly the likelihood of latent class R which is referred to as the cluster around rule $\underset{\sim}{R}$ in this paper. That is,

$$(19) \qquad L(\underset{\sim}{R}) = \prod_{i\in\{R\}} \prod_{j} \left\{ S_j(\theta)^{u_j} (1 - S_j(\theta))^{1-u_j} \right\} = \prod_{i\in\{R\}} \prod_{j} \left[P_j(\theta)^{x_j} (1 - P_j(\theta))^{1-x_j} \right].$$

Suppose a sample {IRT} that fits the IRT model well contains K + 2 latent classes or the clusters around K rules besides Rule $\underset{\sim}{0}$ and Rule $\underset{\sim}{1}$, denoted by $\{\underset{\sim}{0}\}, \{\underset{\sim}{R}_1\}, \{\underset{\sim}{R}_2\} \ldots \{\underset{\sim}{R}_K\}, \{\underset{\sim}{1}\}$, then the IRT sample must be the union set of K + 2 latent classes of R,

$$(20) \qquad \{IRT\} = \{\underset{\sim}{0}\} \cup \{\underset{\sim}{R}_1\} \cup \{\underset{\sim}{R}_2\} \ldots \cup \{\underset{\sim}{R}_K\} \cup \{\underset{\sim}{1}\} = \bigcup_{k=0}^{K+1} \{\underset{\sim}{R}_k\}.$$

By avoiding to count subjects who may belong to the gray area between two clusters, and taking memberships in the clusters to be mutually exclusive, we conjecture that the likelihood of IRT model can be given by the equation below.

$$(21)\quad L(IRT) = \prod_{\substack{K+1 \\ i \in \bigcup_{k=0} \{R_k\}}} \prod_{j=1}^{n} \left\{ P_j(\theta)^{x_j} (1 - P_j(\theta))^{1-x_j} \right\} = \prod_{i=1}^{N} \prod_{j=1}^{n} \left\{ P_j(\theta)^{x_j} (1 - P_j(\theta))^{1-x_j} \right\}.$$

An assumption required in IRT models is the unidimensionality of data. Therefore the union of K + 2 sets expressed in (20) must satisfy the unidimensionality condition in order to yield estimates of logistic parameters. The most intuitive explanation for the union of K + 2 clusters to become unidimensional is that their centroids are located on the principal axis of the IRT sample and the first eigenvalue is considerably larger than the others. Moreover, each class $\{R_k\}$ follows the compound binomial distribution given by Equation 7 with the slip probabilities given in Equation 9. By rewriting the bug distribution in the form of a conventional conditional density function using the relation given in Equation 19, each rule R_k is seen to have the likelihood $\prod_{j=1}^{n} P_j(\theta_R)^{r_j} (1 - P_j(\theta))^{1-r_j}$, which indicates the frequency with which rule R_k is chosen by students, and is hence a measure of the popularity of rule R_k among students. Let us denote the stability and consistency of rule R_k by S_{R_k}, and C_{R_k}, respectively; then each rule R_k is characterized by the values of its stability S_{R_k} and consistency C_{R_k}.

The conditional density function of a given latent class R and the bug distribution of Rule R.

In the previous section, it was shown that IRT could be regarded as a special case of latent classes, and that the conditional probability function, $P_j(\theta)$, of IRT model becomes identical to the bug distribution of Rule $\underset{\sim}{Q}$. The relation was given by Equations 13 and 14. Moreover, the likelihood function of any rule R, written in terms of slip variable u_j and slip probability $S_j(\theta)$ in Equation 15, can be rewritten by using the score x_j, of item j and the IRT conditional probability function $p_j(\theta)$ in Equation 18. The slip variable u_j can be transformed into the item score variable x_j by a relation parallel to Equation 2a, viz.:

$$(22) \qquad x_j = r_j(1 - u_j) + (1 - r_j)u_j = |r_j - u_j|$$

Therefore, the total score, $\sum_{j=1}^{n} x_j$ depends on the elements r_j of rule R, and becomes

$$(23) \qquad \sum_{j=1}^{n} x_j = \sum_{j=1}^{n} r_j(1 - u_j) + \sum_{j=1}^{n} (1 - r_j)\, u_j = \sum_{j=1}^{n} r_j - \sum_{r_j=1} u_j + \sum_{r_j=0} u_j$$

that is, the sum of $1 - u_j$ over the items for which $r_j = 1$ plus the sum of u_j over the items which $r_j = 0$ or, equivalently, $\sum_{i=1}^{n} r_j$, the total score of R plus the difference between the number of slips $0 \rightarrow 1$ and $1 \rightarrow 0$. By taking the expectation of x_j and u_j of Equation 22 for a given θ, Equation 24 is obtained.

$$(24) \qquad \underset{k|j}{\epsilon}(x_k = 1|\theta) = R_j - R_j \underset{k|j}{\epsilon}(u_k = 1|\theta) + (1 - R_j) \underset{k|j}{\epsilon}(u_k = 1|\theta).$$

Replacing $\underset{k|j}{\epsilon}(x_k = 1|\theta)$ by $P_j(\theta)$ and $\underset{k|j}{\epsilon}(u_k = 1|\theta)$ by $S_j(\theta)$, Equation 24 can be rewritten as $P_j(\theta) = R_j - R_j S_j(\theta) + (1 - R_j)S_j(\theta)$. Summing over j from 1 through n, and expressing $S_j(\theta)$ by $P_j\theta)$ or $Q_j(\theta)$ as appropriate, the following equation is obtained:

$$(25) \qquad \sum_{j=1}^{n} p_j(\theta_R) = X_R - \sum_{R_j=1} Q_j(\theta_R) + \sum_{R_j=0} Q_j(\theta_R)$$

where X_R is the number of ones in $\underset{\sim}{R}$. The expected variance of the total score will be $\sum_{j=1}^{n} p_j(\theta_R)Q_j(\theta_R)$. (Lord, 1980). The variance of $\sum_{r_j=1} u_j$ and $\sum_{r_j=0} u_j$ are $\sum_{r_{i=1}} S_j(\theta_R)(1 - S_j(\theta_R))$ and $\sum_{R_{i=0}} S_j(\theta_R)(1 - S_j(\theta_R))$, respectively. Adding the two sums yields $\sum_{j=1}^{n} S_j(\theta_R)(1 - S_j(\theta_R))$ which is equal to expression (10). Therefore, the relationship between the bug distribution of slips away from rule R and the conditional density functions of latent class R is as summarized below:

I) Both have the same likelihood function as can be seen in Equation 19.

II) Both have the same expected variance given in Equation 10.

III) The expectation of the conditonal density function of latent class R, $\epsilon(x_j = 1|R)$ is the sum of the number of ones in $\underset{\sim}{R}$ and the difference of the expectation of number of slips changing from "0 to 1" and that of "1 to 0".

Interpretability of estimated parameters, factors and clusters

Determination of the number of latent rules or classes that are required in cognitive error diagnosis testing is crucial for successful results. Most statistical and psychological modeling such as cluster

analysis, factor analysis and latent class models are developed for finding several groups into which subjects are assigned so that subjects belonging to the same group are more similar than are subjects belonging to different groups. However, it is not unusual to encounter difficulties in interpreting the estimates for the psychological models, clustered groups or factors. Unlike most psychological models, rule space has been developed by emphasizing the importance of interpretability of statistics estimated from the data.

Tatsuoka (1986, 1987) expressed task attributes involved in n items by a binary matrix (called Attribute x Item matrix in which the element Q_{kj} is 1 or 0, and 1 means that item j requires subtask k and $Q_{kj} = 0$ means that item j does not require subtask k. If students use two different strategies to solve the items, then two matrices are constructed with different sets of task attributes and item task vectors. Figures 1 and 2 show two Attribute x Item matrices based on two distinctly different strategies for solving fraction subtraction

Figures 1 & 2 about here

problems. The first strategy (Method A) is to solve the problems by always converting a mixed number (e.g. 3 1/4) to a simple fraction (e.g. 13/4) and adding or subtracting the two fractions. The second strategy, (Method B) involves separating the whole number part from the fraction part, adding or subtracting the two numbers independently, then combining the answers. Method A requires better computational skills while Method B requires deeper understanding of the number system. The borrowing skill is not required by Method A, while it has an important

role in Method B. As a result, erroneous rules resulting from borrowing skills will not appear in students using Method A, but they will often be observed in those who use Method B. Table 1 shows θ_R, ζ, stability, slip dispersion, likelihood of latent classes by Method A and those by borrowing errors when Method B is used. The interpretation of these classes is given in Appendix I. The values in Table 1 used the

Insert Table 1 about here

two-parameter logistic model obtained from a sample of size N = 543 and the group of Method A users are identified by the rule space diagnostic mechanism by Tatsuoka (1986). The third column contains the number of students classified into each of 18 latent classes, students in each class being diagnosed as having the source of corresponding misconceptions as described in Appendix I. Appendix I lists the interpretation of error types of 18 latent classes. The fourth and fifth columns show the positions of the classes in the rule space. Since ζ is fairly large for class 7 (1.48) this class is unusual, so the probability of observing the class 7 misconception will be small, while class 8 (ζ = -.16) will be observed often. The sixth column is the number of slips each class may expect to have. Class 5 has a mean number of almost 15 slips, or 36.6% of the 40 items. Therefore, class 5 represents a very unstable misconception and as such it may be easier to remediate. Classes 1 and 12 are in a fairly stable state compared to classes 5, 6, 7, 8, and 30. The classes located in the neighborhood of the mean θ value of the group tend to be less stable.

Investigation on the Conjecture

The conjecture described in Equation (20) that an IRT sample is the union set of K + 2 latent classes, or clusters around rules $R_1 \ldots R_K$, $\underset{\sim}{0}$ and $\underset{\sim}{1}$ can be examined by a Monte Carlo study. The procedure for testing this hypothesis is as follows:

Hypothesis: Item parameters of the Method A dataset in Table 1 are equal to the estimated item parameters of the Monte Carlo dataset, generated as described below.

1) Generation of the Monte Carlo data

   a. The 18 classes listed in Table 1 are used in this study. A sample of size N = 1000 was generated as the union of subsamples with sizes proportional to the numbers of students classified into the respective classes in the previous dataset, as shown in the third column of Table 1 (N = 328).

   b. Slip probabilities for each class on the 40 items are computed from the original set of item parameters given in Table 2. The estimates of parameters were calibrated from a larger sample (N = 534), including the Method A dataset of N = 328 as a subset.

2) Method of Analysis

   a. Principal component analysis and varimax rotation were carried out for Method A sample and generated data, and their eignevalues were examined.

   b. IRT parameters of Method A sample were calibrated.

   c. IRT parameters of the generated N = 1000 sample were calibrated.

   d. Mean square error was computed and χ^2 test of disparity between the two sets of item parameters were carried out (Lord, 1980, p. 223).

Results of the analysis, d, indicated that the two sets of estimated item parameter values a and b -- one from the actual sample of N = 308 and the other from the generated dataset of N = 1000 -- do not have significant differences, as shown in Table 2.

Insert Table 2 about here

This implies that the conjecture expressed by Equation 20 is supported.

The analysis a, indicated Method A sample and generated data are unidimensional and their eigenvalues (larger than 1.0) are 20.02, 2.17, 2.03, 1.55, 19.82, and 1.62, respectively.

Discussion

This study demonstrates that IRT and latent-class models are algebraically related and the IRT conditional density functions of the items are expressible in terms of those of latent "null state" class. Bug distributions introduced by Tatsuoka and Tatsuoka (1987) are here formulated by using the notion of slips away from the perfect response patterns $\underset{\sim}{R}$ representing a state of knowledge. The bug distribution follows a compound binomial distribution with slip probabilities $S_j(\theta)$. $S_j(\theta)$ equals $P_j(\theta)$ if $R_j = 0$ and $Q_j(\theta)$ if $R_j = 1$. If a bug distribution is expressed by the conditional probabilites of item scores given θ, then it can be considered as the conditional density function of latent class R. The traditional latent class models require the rather strong assumption of statistical independence among classes and further the latent classes must be mutually exclusive and together exhaustive in order to enable formulation of likelihood functions and estimation of

parameters. Moreover, the underlying foundation of latent class models assumes that each state of knowledge is discrete and hence not mutually transferable from one state to another. However, recent advances in cognitive psychology have shown the learning process is very volatile and students change their hypotheses or theories while their learning is in progress. Therefore, the constraints imposed on the latent class models make it difficult to explain theory changes, or to measure change scores, although the models can explain cognitive states of knowledge fairly well as Paulson's model (1985) does.

Rule space respresentation of response patterns enables us to visualize both the IRT and latent class models in a Cartesian Product space of θ and the value of a linear operator $f(\underset{\sim}{x} ; \theta = (\underset{\sim}{P}(\theta) - \underset{\sim}{x}, \underset{\sim}{P}(\theta) - \underset{\sim}{T}(\theta))$ (Tatsuoka, 1985, 1986, Tatsuoka & Tatsuoka, 1987). In this representation θ plays the role of "glueing two contrasting psychological models, IRT and latent classes in a single two-dimensional vector space. By so doing, conceptualization of latent states of knowledge and a continuum scaling of latent ability θ becomes much easier than thinking in the abstract.

Introduction of bug distributions, instead of the traditional conditional density approach of latent classes have made it easier to derive algebraic relationships between IRT and latent class models and made it possible to develop a general model of rule space, which is an expansion of the two leading psychological models. Consistency and stability of rules, or state of knowledge are introduced in the context of distribution theory in this study. However, validation of these new

notions, characterize each erroneous rule, cognitive error or latent class requires further investigation.

The conjecture raised in this study also requires further investigation. This study showed that if the union set of several latent class samples satisfies the unidimensionality condition then their first eigenvector in principal component analysis becomes collinear with the first eigenvector of the IRT sample. In other words, it is plausible to conjecture that the centroids of several latent class samples are on first principal axis of principal component analysis, then the IRT model will also fit the union of these latent-class samples. Grounds for acceptance of this conjecture were provided only by a Monte Carlo study in this paper. More mathematically rigorous investigation of the topic is needed.

References

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

Glaser, R. (1985). The integration of instruction and testing. Paper presented at the ETS Invitational Conference on the Design of Testing for the 21st Century, The Plaza, New York City.

Greeno, J. G. (1980). Instruction for skill and understanding in mathematical problem solving. Paper presented at the 22nd International Congress of Psychology, Leipzig.

Lazarsfeld, P. F. & Henry, N. W. (1968). Latent structure analysis. Boston: Houghton-Mifflin.

Linn, R. L. (1985). Barriers to new test designs. Paper presented at the ETS Invitational Conference on the Redesign of Testing for the 21st Century.

Lord, F. M. (1980). Application of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Paulson, J. A. (1985). Latent class representation of systematic patterns in test responses (ONR) Research Report). Portland, OR: Portland State University.

Tatsuoka, K. K. (1984). Analysis of errors in fraction addition and subtraction problems. (NIE Final Report). Urbana, IL: University of Illinois, Computer-based Education Research.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. Journal of Educational Statistics, 12(1), 55-73.

Tatsuoka, K. K. (1986). Diagnosing cognitive errors: Statistical pattern classification and recognition approach. Behaviormetrika, 19, 73-86.

Tatsuoka, K. K. (1987). Interpretable item relational structure. Manuscript submitted for publication.

Tatsuoka, K. K. (in press). Toward an Integration of item-response theory and cognitive error diagnoses. In N. Fredericksen, R. L. Glaser, A. M. Lesgold, & M. G. Shafto (Eds.), Diagnostic monitoring of skill and knowledge acquisition.

Tatsuoka, K. K., & Tatsuoka, M. M. (1981). Spotting erroneous rules in signed-number arithmetic by the individual consistency index. (Technical Report 81-4-ONR). Urbana, IL: University of Illinois, Computer-based Education Research Laboratory.

Tatsuoka, K. K., & Tatsuoka, M. M. (in press). Bug distribution and pattern classification. Psychometrika.

VanLehn, K. (1983). Felicity conditions for human skill acquisition: Validity an AI-based theory (Research Report CIS-21-ONR). Palo Alto, CA: Cognitive & Instructional Sciences Group, XEROX Research Center

Table 1
Summary Statistics for Eighteen Latent Classes of Method A (Determined from Figure 1) N = 328

	Class	Numbers Classified N	Performance Level θ	Atypicality of class ζ	Mean Slips μ_R	Stability $\mu_R/40$	Variance $\Sigma P_j Q_j$	Stability 0→1 $\sum_{r_j=1} Q_j$	Stability 1→0 $\sum_{r_j=0} P_j$	Consistency $\left[\prod_{j=1}^{40}(1-S_j)\right]^{1/40}$
1	1	29	-2.69	-.80	1.42	.036	.785	.675	.741	.949
2	2	16	-1.22	-.69	5.10	.128	3.085	2.876	2.222	.844
3	3	15	-.75	-.68	7.75	.194	4.294	3.914	3.832	.768
4	4	22	-.46	.75	10.47	.262	4.962	5.712	4.749	.634
5	5	20	.11	.91	14.62	.366	6.813	7.998	6.621	.573
6	6	21	.64	1.74	14.01	.350	6.104	7.384	6.630	.553
7	7	5	-.17	1.48	13.12	.328	5.852	7.395	5.728	.580
8	8	26	.40	-.16	14.16	.354	7.082	7.367	6.796	.599
9	9	6	.60	-.43	11.80	.295	6.379	5.312	6.494	.657
10	10	8	.57	-.24	12.93	.323	6.518	6.605	6.336	.621
11	11	43	.99	.72	6.12	.153	3.303	2.641	3.479	.771
12	12	60	1.19	.86	3.19	.080	2.093	.933	2.252	.881
13	27	11	-.87	-.56	6.74	.169	4.009	3.894	2.844	.802
14	28	10	-1.99	1.01	3.52	.088	1.451	1.700	1.811	.842
15	29	2	-.19	1.53	12.84	.321	5.788	7.617	5.219	.609
16	30	13	-.24	2.74	13.56	.339	5.618	8.631	4.930	.489
17	31	9	-1.18	1.46	6.96	.174	3.182	3.907	3.049	.764
18	32	12	-1.45	.58	5.77	.144	2.476	2.640	3.131	.753

Table 2
Estimated Item Parameters of Sample A and Generated Sample and χ^2 for Testing the Null Hypothesis $a_j = a_j'$, $b_j = b_j'$, $j=1,\ldots,n$.

Method A Sample N=306		Generated Sample N=1000		
a	b	a'	b'	χ^2
1.131	-.034	1.468	.373	.0002
1.504	-.213	2.137	.204	.0003
1.369	.079	1.756	.345	.0002
1.079	-.241	.797	.043	.0001
.725	-.256	.613	-.222	.0000
1.099	-1.220	1.623	-.767	.0002
1.719	.246	1.634	.444	.0002
.693	-1.124	.557	-1.320	.0000
.469	-.734	.340	-.722	.0000
1.889	.201	2.517	.568	.0005
2.136	-.058	2.179	.382	.0003
1.114	-.788	1.312	-.300	.0001
2.377	.473	2.004	.705	.0004
1.451	-.644	1.673	-.355	.0002
1.741	.132	2.248	.425	.0003
1.179	-.581	1.460	-.257	.0001
1.661	.095	2.461	.459	.0004
1.356	-.002	1.688	.392	.0002
2.495	.446	3.979	.746	.0015
1.737	.184	2.457	.519	.0004
1.668	.224	1.792	.477	.0002
1.928	.065	2.541	.264	.0004
2.222	.191	2.262	.347	.0003
.853	-.405	.538	-.187	.0000
1.485	.195	1.424	.271	.0001
1.589	-.964	2.251	-.605	.0004
2.702	.309	2.140	.475	.0003
.715	-1.282	.837	-.929	.0001
.871	-.148	.5605	-.065	.0000
2.371	.322	3.298	.586	.0009
2.090	.094	2.991	.442	.0006
1.053	-.542	1.240	-.254	.0001
3.922	.531	2.253	.930	.0006
1.510	-.542	1.974	-.358	.0002
1.937	.202	2.068	.481	.0004
1.066	-.415	1.263	-.203	.0001
1.727	.120	3.386	.476	.0009
1.604	-.019	2.090	.334	.0003
2.713	.420	3.666	.770	.0014
2.218	.231	3.606	.604	.0011

END

9-87

DTIC