Analysis of Cognitive Structure Using the Linear Logistic Test Model and Quadratic Assignment

María Medina-Díaz

University of Puerto Rico

The cognitive structure of an algebra test was defined and validated using the linear logistic test model (LLTM) and quadratic assignment (QA), respectively. The LLTM is an extension of the Rasch model with a linear constraint that describes the difficulty of a test item in terms of the cognitive operations required to solve it. The cognitive structure of a test is specified using the weight matrix **W**. The cognitive structure defined here was based on a set of eight production rules that represented the mathematical procedures employed in solving linear equations with one variable. A 29-item test was constructed and administered to 235 ninth-graders. Item response data were analyzed using Fischer & Formann's (1972) LLTM computer program. A QA confirmatory approach was used to validate the cognitive structure of the test. The structure was validated—examinees solved the items using the set of rules specified in the **W** matrix. The parameters estimated using the LLTM are quantitative indexes of the difficulties of each of the cognitive rules included in the **W** matrix. *Index terms: componential models, confirmatory analysis, content validity, linear logistic test model, quadratic assignment, cognitive structure, validation.*

The development of psychometric methods for assessing individuals' varied and complex abilities, and the integration of such assessment with learning and instruction, has been a longstanding challenge. Currently there is an interest in accounting for the cognitive processes examinees use in solving a test item, rather than only in evaluating the final answer. One approach is to design test items that elicit a certain type of cognitive information from examinees. To do this, it is necessary to establish a relationship between the item content, the psychometric characteristics of the item, and the cognitive or psychological attributes the item proposes to measure.

Developments in item response theory (IRT) models and techniques are promising for this line of inquiry. IRT methodologies make it feasible to analyze test items using a psychological theory as a framework to show what attributes of the items lead to differences in examinee performance. For example, differences found among items with respect to their underlying psychometric characteristics can also reveal differences in the cognitive skills that examinees use to answer the items. IRT componential models (e.g., Embretson, 1983; Fischer, 1973, 1974; Whitely, 1980) have been used to accomplish this. One of these models, the linear logistic test model (LLTM), was investigated here.

The Linear Logistic Test Model

Description

Fischer (1973, 1974) presented one of the first attempts to account for the cognitive skills—the type of knowledge or procedures required to solve the items—that underlie responses to test items. Fischer (1973) found that the difficulties of a set of calculus differentiation items could be explained by eight basic cognitive operations that the examinee must implement in order to solve a differentiation problem. Based on a hypothetical or a priori structure of the test, he postulated that the item difficulty could be expressed in terms of the contribution from each separate operation.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 17, No. 2, June 1993, pp. 117-130

0146-6216/93/020117-14$1.95

Accordingly, the item difficulty parameter is a linear function of the occurrence and difficulty of the necessary operations for solving the task. The resulting componential model is called the LLTM. Thus, the difficulty of item i is given by

$$\beta_i^* = -\sum_{k=1}^{m} w_{ik}\eta_k + c \ , \quad (1)$$

where

β_i^* is the difficulty parameter for item i ($i = 1, 2, \ldots, n$);

w_{ik} is the weight (occurrence) assigned to operation k that is involved in solving item i;

η_k is the "basic" parameter (Fischer, 1973, p. 361) for operation k ($k = 1, 2, \ldots, m$); and

C is a normalization constant, and is defined as

$$c = -\frac{1}{n}\sum_{i}^{n}\sum_{k}^{m} w_{ik}\eta_k \ , \quad (2)$$

such that $\sum_{i=1}^{n} \beta_i^* = 0$.

Mathematically, the LLTM is an extension of the Rasch model with a linear constraint (Fischer, 1983). This linear constraint (Equation 1) permits the specification of the contributions of every operation defined a priori to the difficulty of the item.

For every item i, a vector of weights is defined as $\mathbf{w}_i = (w_{i1}, w_{i2}, \ldots, w_{im})$; the kth element denotes the occurrence of the kth operation in answering the item. The set of all item vectors can be presented in the form of a weight matrix $\mathbf{W}$ with dimensions $n \times m$ ($m < n$), in which each row contains the vector $\mathbf{w}_i$. The weights can be either 1 or 0, indicating the presence or absence of a given operation in solving an item, respectively. In other instances, the weights can be the frequency with which the operation is involved in solving the item (Spada, 1977; Spada & McGaw, 1985). Once $\mathbf{W}$ is specified and the examinees' responses dichotomously scored, the ηs must be estimated. Conditional maximum likelihood estimates of the ηs can be obtained using Fischer & Formann's (1972) computer algorithm. Using these η estimates, the conditional maximum likelihood estimates of the βs are computed using Equation 1.

Applications

The LLTM has been used in a number of applications (e.g., Embretson & Wetzel, 1987; Fischer, 1973, 1978; Fischer & Formann, 1982; Green & Smith, 1987; Hornke & Habon, 1986; Koponen, 1983; Mitchell, 1983; Sheehan & Mislevy, 1990; Spada, 1977; Whitely & Schneider, 1981). Despite the success of this model with ability tests, the application to achievement testing has been limited. Further, the operations involved in solving items have been defined broadly. For example, Fischer (1973) defined the hypothesized operations by the item topics; Spada (1977) defined the hypothesized operations by the properties or features of the task. Neither provided a detailed account of the specific cognitive processes underlying the operations.

The way in which these operations have been defined and specified for test construction purposes is an area of concern in light of the recent conceptualizations of knowledge and learning developed from cognitive psychology research (e.g., Anderson, 1985; Rumelhart & Norman, 1978). The recent evaluation of document literacy skills by Sheehan & Mislevy (1990) using the LLTM constitutes an advance in this regard. They examined these skills through parsing models (i.e., semantic representations) that represent the individual's information-processing mechanisms in reading text. A more thorough analysis of the examinees' knowledge can be conducted using semantic networks or production rules (PRs) representing an individual's knowledge.

Anderson (1985), for example, distinguished between declarative knowledge and procedural knowledge. Declarative knowledge is organized into systems of information (e.g., semantic networks) that give meaning to objects and events that are perceived. Procedural knowledge, in contrast, is knowledge about how to perform or do various operations. For instance, problem-solving performance can be represented as the serial application of a set of PRs. A PR is a two-part statement (condition–action) that specifies a given action to be executed under specific conditions. When the condition (the IF side) is true, a specific action is taken (the THEN side). PRs can be linked to production systems—that is, whenever the action is taken by one PR this activates the execution of another PR (see Anderson, 1982; Sleeman, 1984). The result of applying a PR is a transformation of information, such as a sequence of steps in solving a problem. In the present study, PRs represented the kind of specific knowledge employed in solving linear equations. Operations referred to the set of PRs presumed to be executed by the examinee in solving a given equation.

Related to these ways of representing knowledge are the methods employed by cognitive psychology to validate the notions about the processes that underlie learning and performance. For instance, verbal reports are valid empirical methods for collecting information about mental processes under certain conditions and assumptions (Ericsson & Simon, 1980). In this technique, called protocol analysis, examinees report what they are thinking while performing a task. Protocol analysis has been used as a major research tool in artificial intelligence (e.g., Newell & Simon, 1972) as well as in mathematics education (e.g., Resnick & Ford, 1981; Schoenfeld, 1985). Thus, if an item is taken as a single task, the process of solving it can be modeled by a set of operations based on protocol data. This approach was used here—the steps performed by the examinees in solving linear equations were translated to PRs using a protocol analysis.

Finally, another critical issue related to the application of the LLTM to various measurement situations is the scarcity of approaches for evaluating the validity of **W**. Statistical procedures such as the likelihood ratio are used to test the overall validity of the LLTM against other models (e.g., the Rasch model) or for comparing alternative models (e.g., Fischer, 1973, 1974; Fischer & Formann, 1982; Whitely & Schneider, 1981). Inferences are made in terms of the adequacy of the model in describing the observed data.

The LLTM relies on the specification of **W**. However, no other statistical procedures other than the likelihood ratio test have been available to confirm the structure of this matrix and, therefore, to support the researcher's initial hypothesis about the cognitive complexity of the test items. If the **W** matrix does not correspond to the way in which the examinees actually solve the items, the estimates of the ηs will not convey valid information about their responses (Fischer, 1974, 1977). Thus, the validation of the **W** matrix is an important aspect of using the LLTM for both test analysis and test design. The quadratic assignment (QA) paradigm was used here to confirm the proposed cognitive structure of the test.

Quadratic Assignment

QA techniques have been used to verify theoretical structures of observed data in education, psychology, geography, and economics (e.g., Baker & Hubert, 1977; Hubert, Golledge, & Costanzo, 1981; Hubert, Golledge, Costanzo, Gale, & Halperin, 1984; Hubert & Schultz, 1976). The QA paradigm refers to a group of data analysis techniques dealing with object orderings and with the relationship of the proximity matrix—denoted by **Q**—to a hypothetical underlying organization of those objects—referred to as the structure matrix **C**. Hubert & Schultz (1976) described two types of QA analyses: confirmatory and exploratory.

The exploratory approach identifies a particular type of structure in the data without an a priori hypothesis about this structure. The confirmatory approach evaluates whether or not a given hypothetical structure can be used to explain the pattern posited in the actual proximity data. QA confirmatory analysis was used here because an a priori hypothesis existed about how the data should be organized. Hubert & Schultz (1976) provided a discussion of the statistical underpinnings of QA as well as some applications in education and psychology. Also, Baker & Hubert (1977) presented an introduction to QA techniques and didactic examples in the field of educational psychology.

To perform QA analyses it is necessary to have some empirical measure of proximity between the pair of objects under study. **Q** contains the proximity measures, such as a similarity measure between test items. In this study, the elements of **Q** indicated the extent to which a given pair of items was solved using the same combination of PRs. **Q** contained empirical information based on examinees' responses to every pair of items.

The structure matrix **C** constitutes the researcher's a priori hypothesis about the organization of the similarity data contained in **Q**. Here, **C** was obtained from the weight matrix **W** defined under the LLTM. The elements of **C** contained the hypothetical number of PRs that a given pair of items had in common. Hence, **C** conveyed the hypothetical cognitive structure of the test.

QA confirmatory analysis tests whether the specific structure of **C** is represented in **Q**. The null hypothesis is that the rows and columns of **Q** are assigned at random relative to the structure specified in **C** (Baker & Hubert, 1977). An index Γ is used to measure the strength of the association between **Q** and **C**. Γ is the sum of all cross-products of the corresponding elements in **Q** and **C**,

$$\Gamma = \sum_{x}^{n} \sum_{y}^{n} [q(x,y)][c(x,y)] \ , \quad (3)$$

where x and y are two items, ($x \neq y$), and $q(x,y)$ and $c(x,y)$ are the elements (row x, column y) for these items in **Q** and **C**, respectively. Because the diagonal elements of **Q** and **C** are 0s, this sum is actually taken over the $n(n-1)/2$ off-diagonal elements of the two matrices (Baker & Hubert, 1977). Hubert & Schultz (1976) provided the mathematical derivation of Γ.

QA confirmatory analysis requires that the observed value of Γ be compared with a randomization distribution for this index. **Q** establishes the particular ordering for the test and generates the observed value of the index Γ. The null hypothesis is used to obtain a reference distribution for Γ using a randomization procedure. In theory, given **C**, a permutation procedure based on $n!$ possible orderings of the rows and columns of **Q** can produce a reference distribution for Γ. Each ordering produces a Γ value, and when these values are tabled through complete enumeration an exact distribution of Γ is obtained (Baker & Hubert, 1977). Alternative approximations to the exact randomization distributions, such as the Z unit normal, can be used to test the null hypothesis (Anselin, 1986).

To test the null hypothesis, the observed value of Γ and the critical value derived from the reference distribution are compared. If the observed value of Γ falls in the critical region (e.g., $p < .05$), the null hypothesis of random allocation is rejected indicating that **Q** reflects the hypothetical structure of the test represented in **C**.

Method

Purpose

The goals of the present study were (1) to use the LLTM to define the cognitive structure of an algebra test involving linear equations with one variable, and (2) to use the QA confirmatory approach to validate the hypothetical cognitive structure of the test.

Development and Administration of the Algebra Test

The first task was to construct the items of an algebra test. Five textbook series (Brumfiel, Golden, & Heins, 1986; Dolciani, Brown, & Cole, 1988; Dolciani, Sorgenfrey, & Graham, 1988; Foster, Rath, & Winters, 1990; Price, Rath, & Leschensky, 1989) were consulted regarding the skills included in pre-algebra and algebra courses. Based on this review, 15 items covering linear equations were constructed and administered to 10 high school students using protocol analysis (i.e., the examinees were asked to think aloud while solving the equations).

Eight PRs were inferred by using the skills covered in the textbooks, the protocol data of the 10 students, and their actual problem solutions. These PRs—denoted P1, P2, P3, P4, P5, P6, P7, and P8—are described in Figure 1. It was assumed that each PR modeled part of the procedural knowledge required to solve linear equations with

Figure 1
Operations and PRs Relevant for the Solution of Linear Equations With One Variable

Operation 1: Collecting

P1 IF more than one numeral exists on one side of the equation and these numerals are related by addition (or subtraction), THEN write the total (or difference) of these numbers.

P2 IF there is one numeral and a term on one side of the equation or there are two numbers related by multiplication, THEN write the product of those terms or numbers.

P3 If more than one like term exists on one side of the equation and those terms are related by addition (or subtraction), THEN write the total (or difference) of those terms.

Operation 2: Balancing

P4 IF there is only one term in the equation and the term and a numeral exist on the same side of the equation and are related by addition or subtraction, THEN add the additive inverse of the numeral to both sides of the equation.

P5 IF one like term exists on each side of the equation and the terms are related to numerals by addition (or subtraction), THEN subtract (or add) the like term on one side of the equation from itself and from the like term on the other side of the equation.

Operation 3: Removing Parentheses

P6 IF like terms or numerals exist inside the parentheses related by any arithmetic operation, THEN execute this operation and write its result.

P7 IF the elements inside the parentheses are not like terms and they are related by addition or subtraction and one term or numeral exists outside of the parentheses, THEN multiply out the term or numeral by each one of the elements inside the parentheses.

Operation 4: Solve for the Variable

P8 IF the goal is to evaluate a variable at one side of the equation and the variable is related to a numeral by multiplication, THEN divide by the numeral on both sides of the equation and the result is "variable = quotient."

one variable (Gagné, 1985; Lewis, 1981; Mayer, 1980; Sleeman, 1984).

A 29-item test was constructed according to these PRs. Each item required the use of a particular combination of PRs to correctly solve the equation. 14 of the possible combinations were represented in the 29 items of the test. Table 1 shows the PRs needed to solve each item. Each item required from two to six PRs applied in a distinct sequence (e.g., for Item 7 P1 must be performed before P8). It was assumed that these combinations varied in level of difficulty depending on the number of PRs included. Two experts (an algebra teacher and a mathematics education professor) and a graduate student in mathematics education evaluated the test to determine if the items assessed the skills covered in the mathematics curriculum for ninth grade.

The test was administered to 235 ninth graders enrolled in introductory algebra classes at two high schools during April, 1991. Students were asked to show all work. The administration time was approximately 50 minutes, equivalent to a class period.

Finally, the data were examined to ensure that the Rasch model assumptions of no guessing and common discrimination were met. In order to evaluate the assumption of no guessing, the item response functions under the three-parameter model using the low-scoring examinees and the total group were inspected (Mislevy & Bock, 1989). The likelihood ratio test was employed to test the assumption of common item discrimination parameters of the test items (Andersen, 1973; Gustafsson, 1980).

Table 1
Number and Sequence of PRs Required to Solve the Items of the Algebra Test

Number and Sequence of PRs	Item Numbers
Total PRs: 2	
P1, P8	7
P2, P8	2
P1, P4	3
P3, P8	5, 15
Total PRs: 3	
P5, P3, P8	1, 10
P4, P1, P8	4, 6, 8, 13
P6, P2, P8	9, 19, 21, 24
Total PRs: 4	
P3, P4, P1, P8	12, 16
P5, P3, P4, P8	14
P5, P3, P4, P1	17
P7, P5, P3, P8	20, 22, 23, 25
Total PRs: 5	
P5, P3, P4, P1, P8	11
P6, P2, P5, P3, P8	18, 27, 29
Total PRs: 6	
P7, P5, P3, P4, P1, P8	26, 28

LLTM Test Analysis

The weight matrix **W** and the examinees' dichotomous item responses were used as input for the LLTM computer program (Fischer & Formann, 1972). Table 2 shows the 29 $\times$ 8 weight matrix **W**, which contains the vectors of PRs required for solving each of the 29 items. Each element in these vectors indicates the occurrence (1) or nonoccurrence (0) of a given PR, k, in solving an item. Conditional maximum likelihood estimates of the η and β^* parameters were obtained in this phase.

The LLTM program was used a second time with **W** defined as an identity matrix. This format reflects the Rasch model structure as a special case of the LLTM in which each item defines a separate η_k and C is equal to 0 (Equation 1). Conditional maximum likelihood estimates of the βs also were calculated under this model. The two sets of βs—those from the LLTM and from the Rasch model—were compared using the conditional likelihood ratio test (Andersen, 1973, 1980; Fischer, 1973, 1974). This test indicates whether the use of the LLTM results in different β estimates than those under the Rasch model. If so, it provides justification for using the LLTM rather than the simpler Rasch model.

Table 2
Items and the **W** Matrix of the 29-Item Test

	PR							
Item	1	2	3	4	5	6	7	8
1. $4x = 2x - 1$	0	0	1	0	1	0	0	1
2. $19y = (8)(3)$	0	1	0	0	0	0	0	1
3. $12 + c = 4$	1	0	0	1	0	0	0	0
4. $7 + 14p = 7$	1	0	0	1	0	0	0	1
5. $20a - 9a = 8$	0	0	1	0	0	0	0	1
6. $15m + 7 = 6$	1	0	0	1	0	0	0	1
7. $5 + 4 = 5a$	1	0	0	0	0	0	0	1
8. $2y - 3 = 11$	1	0	0	1	0	0	0	1
9. $(15 - 9)3x = 1$	0	1	0	0	0	1	0	1
10. $7a = -11a + 7$	0	0	1	0	1	0	0	1
11. $20 + 10x = 3 - 5x$	1	0	1	1	1	0	0	1
12. $15a - 4a + 6 = 9$	1	0	1	1	0	0	0	1
13. $5 + 3x - 2 = 7$	1	0	0	1	0	0	0	1
14. $3e - 10 = 5e$	0	0	1	1	1	0	0	1
15. $-1 = 9k - 6k$	0	0	1	0	0	0	0	1
16. $17d + 13d - 9 = 10$	1	0	1	1	0	0	0	1
17. $9n - 3 = 8n + 6$	1	0	1	1	1	0	0	0
18. $7e(11 - 8) = 7e + 3$	0	1	1	0	1	1	0	1
19. $6(6y + 3y) = 0$	0	1	0	0	0	1	0	1
20. $16x = 7(7x + 3)$	0	0	1	0	1	0	1	1
21. $m(4 - 1) = 10$	0	1	0	0	0	1	0	1
22. $15p = 5(2p - 4)$	0	0	1	0	1	0	1	1
23. $7(3c - 5) = 20c$	0	0	1	1	1	0	1	0
24. $9 = 7(4c + c)$	0	1	0	0	0	1	0	1
25. $3y = (9y - 8)3$	0	0	1	0	1	0	1	1
26. $9(2 + 5x) = 8x + 2$	1	0	1	1	1	0	1	1
27. $9(3p + 4p) = 7 - 7p$	0	1	1	0	1	1	0	1
28. $(7y - 4)4 = 5 + 8y$	1	0	1	1	1	0	1	1
29. $1 - 3k = 5(17k - 8k)$	0	1	1	0	1	1	0	1

Quadratic Assignment

QA procedures were used to determine whether the hypothetical cognitive structure of the test (**C** matrix) was present in the examinees' actual item solutions (**Q** matrix).

Defining the **Q** *matrix.* In general, the elements of **Q** represent a relationship of similarity between every pair of items (x, y) of the test $(x, y = 1, 2, \ldots, n; x \neq y)$. **Q** is a $n \times n$ matrix in which both the xth row and the xth column refer to item x, and the entry in row x and column y is denoted by $q(x, y)$. It is also assumed that the diagonal elements are 0, $q(x,x) = 0$. Here the elements of **Q** indicated the extent to which the same PRs were used by the examinees in solving a pair of test items.

The elements of **Q** were computed in two stages. First, because the examinees' actual item responses were considered in the definition of **Q**, the vectors of item responses for each examinee were inspected in order to determine what PRs the examinees actually used. For each examinee, every step of the process of solving an equation was examined and compared with the vector of PRs proposed for solving that particular item (Table 2). If an examinee used a given PR, a value of 1 was assigned ($d_{kx} = 1$) to that rule for that item. A particular PR was judged as being used in solving the item only when the examinee's written work in the test sheets showed its execution. A value of 0 ($d_{kx} = 0$) was assigned when the PR was not used. Therefore, the selection variables (i.e., the d values) described the examinee's use or nonuse of each of the eight PRs in solving an item. A computer program stored the vectors of PRs that each of the 235 examinees used to solve the 29 items.

Secondly, the **Q** matrix elements were calculated. At this stage, for every examinee who solved a given item there was a vector with 1s and 0s corresponding to whether the PRs were used or not used by the examinee to solve that item. For every pair of items (x,y where $x \neq y$) solved by an examinee, the cross-product of the selection variables associated with each kth PR used was computed ($d_{kx}d_{ky}$) and multiplied by the corresponding η_k. The ηs of the eight PRs were obtained from the LLTM program. The result was a vector of m values $(d_{kx}d_{ky})\eta_k$ for that pair of items. These m values were added, and the total represented the measure of similarity between these two items. This procedure was repeated for every pair of test items solved by each examinee.

As a result, a given element $q(x,y)$ of **Q** corresponding to the items x, y ($x \neq y$) was defined as

$$q(x,y) = \sum_{j=1}^{N} \sum_{k=1}^{m} (d_{kx}d_{ky})\eta_k \ , \tag{4}$$

where

d_{kx} and d_{ky} are selection variables (1,0) associated with each kth PR,

η_k is the estimate of the kth "basic" parameter,

m denotes the number of ηs ($k = 1, 2, \ldots, 8$), and

N denotes the number of examinees ($j = 1, 2, \ldots, 235$).

Q was a 29 × 29 symmetrical matrix and its elements summarized the 235 examinees' patterns of responses to each pair of items as a function of the eight ηs.

Defining the structure matrix, **C**. The structure matrix **C** was also a 29 × 29 symmetrical matrix. Its elements, denoted by $c(x,y)$, contained the sum of the cross-products of the weight vectors of the **W** matrix. Thus, the elements were computed as follows:

$$c(x,y) = \sum_{k=1}^{m} w_{kx}w_{ky} \ , \tag{5}$$

where w_{kx} and w_{ky} are the weights of the kth PR in items x and y, respectively. Also, the diagonal elements of **C** were set to 0. In **C**, for every pair of distinct items, the weights of the eight PRs were cross-multiplied. Because these weights were 1s and 0s, the elements of **C** reflected the common number of PRs required to solve every pair of items. The elements of **C** are measures of the similarity between every pair of test items. Therefore, **C** postulates a hypothetical relationship between the test items based on the information of **W**.

QA analysis of the **Q** *and* **C** *matrices.* Four QA cases were analyzed to evaluate the correspondence between the patterns of item responses observed in various datasets (defined as Cases 1–4 below) and the cognitive structure of the test. Case 1 used the item responses of the total sample of examinees ($\mathbf{Q}_{\text{total}}$). Case 2 was based on the item responses of the low-scoring examinees ($\mathbf{Q}_{\text{low}}$); Case 3 was based on the item responses of the high-scoring examinees ($\mathbf{Q}_{\text{high}}$). In Case 4, the empirical pattern of responses of both subgroups was compared ($\mathbf{Q}_{\text{low}}$ vs. $\mathbf{Q}_{\text{high}}$).

Of the 235 examinees, 12 examinees answered all items correctly. The test mean was 23 with a standard deviation (SD) of 5.38. 94 examinees scored below the mean; these examinees constituted the low-scoring subgroup of examinees. The mean score of this subgroup was 17 with a SD of 5.22. The high-scoring subgroup included 141 examinees; their mean score was 27 and the SD was 1.48.

For Case 1, $\mathbf{Q}_{\text{total}}$ was compared to the **C** matrix. (The item responses of 235 examinees were used in the QA analyses, but data from 223 of the 235 examinees were employed in the estimation of the ηs). The elements of $\mathbf{Q}_{\text{total}}$ were computed as described above (Equation 4). The elements of **C** were defined as in Equation 5.

For Case 2, $\mathbf{Q}_{\text{low}}$ was compared to the **C** matrix. The elements of $\mathbf{Q}_{\text{low}}$ were defined as in Equation 4, but the sum was taken over the 94 examinees of the low-scoring subgroup. **C** was the original structure matrix, and its elements were defined as in Equation 5. A similar procedure was used for Case 3, but the sum was taken over the 141 examinees in the high-scoring subgroup.

For Case 4, the low-scoring subgroup ($\mathbf{Q}_{\text{low}}$) was compared to the high-scoring subgroup. $\mathbf{Q}_{\text{low}}$ was defined as in Case 2, and the **C** matrix was $\mathbf{Q}_{\text{high}}$, as defined in Case 3. Thus, the **C** matrix contained empirical information rather than a hypothetical structure as in the previous QA analyses. This was done to determine if the responses of the examinees in the low-scoring subgroup presented patterns similar to those of the high-scoring examinees, here defined as **C**.

Anselin's (1986) MICROQAP computer program was used to perform the QA analyses. Using the various **Q** and **C** matrices as the input data, the MICROQAP program executed a series of random permutations (up to 1,000), constructing an empirical frequency distribution for the Γ test statistic. In addition, MICROQAP computed the Z unit normal and the Pearson Gamma III distributions as reference distributions. In each of the QA cases, the probability values associated with the Z distribution were used for determining the degree of agreement between the elements of the various **Q** and **C** matrices. If the probability values were less than the level of significance ($p < .05$), the null hypothesis that the rows and columns of **Q** were assigned at random relative to the structure specified in **C** was rejected, and it was concluded that the **Q** matrix reflected the hypothetical structure of the **C** matrix.

Results

Evaluating the Rasch Model Assumptions

The item response functions of the test items under the IRT three-parameter model using the low-scoring examinees and the total sample showed low guessing parameter values using both datasets. Seven items had nonzero guessing parameter estimates. The average guessing parameter estimate of these items was approximately .12, with SD = .19. Due to content considerations, these seven items were not eliminated from the test. It was concluded that the minimum guessing assumption was met.

The likelihood ratio test indicated nonsignificant differences in the item discrimination parameters calculated using the low-scoring and high-scoring examinees, and the total sample ($\chi^2_{(28)} = 40.37, p > .05$). This implies that the assumption of common item discrimination parameters held; therefore, the 29-item test data can be assumed to fit the Rasch model.

Estimates of the ηs

Table 3 shows the estimates of the eight ηs and

Table 3
$\hat{\eta}$ for the Eight PRs and Their Standard Errors (SEs)

PR	$\hat{\eta}$	SE
P1	-.308	.138
P2	.477	.179
P3	-.253	.134
P4	-.300	.138
P5	-.516	.125
P6	-1.710	.164
P7	-.754	.104
P8	-.921	.146
Mean	-.536	.102
SD	.590	.066

their standard errors. These estimates are expressed on the Rasch model original metric (Fischer, 1977, 1983). The mean of the η estimates was $-.54$ with SD = .59.

Seven of eight η estimates were negative, ranging from $-.253$ to -1.710, with the latter occurring for P6, which deals with execution of arithmetic operations inside of parentheses. Only P2, which refers to the multiplication of similar elements, obtained a positive estimate ($\eta = .477$) and, therefore, had the highest difficulty value of the ηs. The standard errors of the ηs ranged from .104 to .179 indicating an acceptable precision of the estimates.

The η estimates indicated that the PRs contributed differently to the β^*. Because the solution of a test item involved a particular combination of PRs as shown in the **W** matrix (Table 2), the difficulty of the items depended on the sum of the ηs associated with the PRs (Equation 1). As expected, a general pattern was found: Item difficulty increased as a function of the number of PRs required to solve the item. However, there was some overlap between the difficulty estimates associated with the items that included three and four PRs.

On the other hand, Table 4 shows that the conditional likelihood ratio test between the βs obtained under the LLTM and the Rasch model yielded significant results ($\chi^2_{(28)} = 209.26$, $p < .05$). This indicates that the βs obtained through the LLTM were significantly different than those computed using the Rasch model. Because the likelihood ratio test gives an overall goodness-of-fit measure, this implies that the LLTM accounted for the item difficulties differently than did the Rasch model.

Table 4
LLTM β^* and Rasch Model β Item Difficulty Estimates and Their Differences ($\beta - \beta^*$)

Item	β^*	β	$\beta - \beta^*$
1	-.26651	-.20288	.064
2	-1.51144	-1.32716	.184
3	-1.34766	-1.06611	.282
4	-.42697	.82293	1.250
5	-.78204	-.94890	-.167
6	-.42697	-.28286	.144
7	-.72656	-1.63735	-.911
8	-.42697	-1.39928	-.972
9	.19878	.87687	.678
10	-.26651	-.24252	.024
11	.34134	.68436	.343
12	-.17419	-.36581	-.192
13	-.42697	-.40850	.018
14	.03308	-.49654	-.530
15	-.78204	-.40850	.374
16	-.17419	-.12561	.049
17	-.57935	-.73476	-.155
18	.96709	.93002	-.037
19	.92818	1.05982	.132
20	.48727	1.49851	1.011
21	.19878	-.78606	-.985
22	.48727	-.45206	-.939
23	-.13383	-.08791	.046
24	.19878	.02176	-.177
25	.48727	.56919	.082
26	1.09512	1.40416	.309
27	.96709	1.23508	.268
28	1.09512	.68436	-.411
29	.96709	1.18569	.219
Maximum	1.095	1.499	1.250
Minimum	-1.511	-1.637	-.985
Mean	0.000	0.000	0.000
SD	.711	.898	.531

Quadratic Assignment

The descriptive statistics of the **Q** matrices and the **C** matrix are shown in Table 5.

Case 1—$\mathbf{Q}_{\text{total}}$ *vs.* **C**. The elements of $\mathbf{Q}_{\text{total}}$ are shown in the upper triangle of Table 6. All elements of $\mathbf{Q}_{\text{total}}$ were negative. This was a direct effect of the fact that seven of the eight ηs were negative. The size of the $\mathbf{Q}_{\text{total}}$ elements depended on the contribution of the examinees' responses based on the PRs and of the ηs associated with these PRs (Equation 4). The elements ranged from -28 to -386 with a mean of -202.2 and SD of 74.13.

The structure matrix **C** representing the hypothetical cognitive structure of the test is shown in the lower triangle of Table 6. The mean was 1.91 with SD = 1.13. The values of the elements in **C** ranged from 0 to 6, indicating the number of PRs required in solving every pair of test items. For instance, a value of 0 meant that

Table 5
Mean and SD of the **Q** and **C** Matrices and Number of Examinees (N) Used in the Computation of the **Q** Matrices

Statistic	$\mathbf{Q}_{total}$	$\mathbf{Q}_{low}$	$\mathbf{Q}_{high}$	**C**
N	235	94	141	
Mean	−202.20	−65.33	−136.90	1.91
SD	74.13	24.50	50.57	1.13

the two items were solved using different PRs—that is, there were no common PRs between those items. A value of six indicated that the solution to the two items involved six common PRs.

The initial ordering of $\mathbf{Q}_{total}$ when compared with **C** yielded an observed Γ value of −352,326. This value exceeded the largest value of the Γ in the sample of 1,000 random permutations ($p < .05$). The normalized value of Γ was −4.99, which was significant using the Z unit normal distribution as a reference distribution ($p < .05$). The correlation between $\mathbf{Q}_{total}$ and **C** (provided by MICROQAP) also was significant ($r = -.58$, $p < .05$). These results indicated that the null hypothesis of random matching of the two matrices was rejected. Thus, the alternative hypothesis of matching based on a substantive relationship between these matrices was accepted.

Case 2—$\mathbf{Q}_{low}$ *vs.* **C.** The elements of $\mathbf{Q}_{low}$ ranged from −7.9 to −126.3; the mean was −65.33 with SD = 24.5. The initial ordering of $\mathbf{Q}_{low}$ when compared with **C** yielded an observed Γ value of −113,032, and a normalized Γ of −4.66. The correlation between the matrices was −.53. The probability values associated with the reference distributions of Γ based on the randomization procedure and the Z distribution were .01 and 0.00, respectively, permitting rejection of the null hypothesis of random matching between the elements of $\mathbf{Q}_{low}$ and the original structure matrix **C** ($p = .05$).

Case 3—$\mathbf{Q}_{high}$ *vs.* **C.** The mean of the $\mathbf{Q}_{high}$ matrix elements was −136.9 with SD of 50.57. The elements ranged from −18.4 to −287.1. The QA results showed an observed Γ index of −239,294 and a correlation between $\mathbf{Q}_{high}$ and **C** of −.59. This observed value of Γ again fell in the critical region of the Z distribution. Thus, the null hypothesis of random matching between both matrices also was rejected at $p = .05$.

Case 4—$\mathbf{Q}_{low}$ *vs.* **C** *(defined as* $\mathbf{Q}_{high}$*).* An observed association index Γ of 8,211,410 was obtained between $\mathbf{Q}_{low}$ and the **C** matrix based on $\mathbf{Q}_{high}$. This was the largest Γ value obtained and was due to the large element values of the two matrices. The correlation between $\mathbf{Q}_{low}$ and **C** here was .94, which was also the highest correlation obtained. The probability values associated with the reference distributions of the Γ index showed values close to 0.00, and hence were significant at $p = .05$; the null hypothesis of random matching between the two matrices also was rejected.

Discussion

The significant agreement between $\mathbf{Q}_{total}$ and **C** (Case 1) suggests that the patterns of item solutions of the total sample of examinees reflected the hypothetical structure in the **C** matrix. This was also true for the low- and high-scoring subgroups. The cognitive structure of the test represented by **C** was reflected in how both the low- and high-scoring examinees solved the test items. Therefore, the correspondence between **Q** and **C** indicated that the examinees actually used the operations defined under the LLTM framework.

Because the elements of **C** were based on information in **W**, and **Q** reflected the structure of **C**, it can be concluded that **Q** also reflected the structure of **W**. The cognitive structure of the test was validated using QA procedures; that is, the examinees actually used the hypothesized cognitive rules. Consequently, **W** described the operations required to solve the test items, and the ηs based on this structure represented these operations.

Because of the nature of the skills involved in solving the equations, it was expected that there would be some correspondence between **Q** and **C**. There was a set of eight specific rules to be used, and the evaluation of the examinees' performance was contingent on these rules. In the

Table 6
Q Matrix Based on the Total Sample (Upper Triangle) and C Matrix (Lower Triangle)

	Item																												
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	0	-194	-33	-158	-241	-194	-193	-196	-185	-263	-221	-214	-193	-218	-238	-215	-59	-223	-187	-247	-208	-253	-148	-222	-239	-219	-247	-218	-244
2	1	0	-32	-162	-198	-200	-203	-205	-124	-199	-194	-200	-201	-202	-195	-200	-28	-135	-134	-188	-153	-198	-107	-173	-190	-190	-176	-190	-162
3	0	0	0	-95	-31	-101	-88	-104	-30	-34	-96	-96	-102	-54	-32	-94	-65	-32	-27	-33	-31	-31	-31	-32	-31	-89	-33	-88	-30
4	1	1	2	0	-158	-228	-218	-235	-146	-161	-220	-224	-229	-178	-158	-224	-84	-140	-146	-151	-152	-159	-96	-160	-155	-210	-157	-209	-152
5	2	1	0	1	0	-196	-197	-198	-187	-247	-206	-219	-195	-206	-243	-219	-46	-204	-191	-231	-209	-242	-134	-226	-229	-203	-234	-205	-230
6	1	1	2	3	1	0	-263	-278	-175	-200	-258	-269	-269	-210	-193	-266	-94	-170	-168	-189	-185	-194	-116	-198	-189	-247	-195	-249	-188
7	1	1	1	2	1	2	0	-267	-178	-200	-252	-262	-261	-205	-193	-261	-84	-170	-173	-186	-187	-196	-107	-199	-188	-241	-195	-241	-185
8	1	1	2	3	1	3	2	0	-180	-202	-264	-274	-279	-222	-196	-271	-96	-174	-176	-192	-191	-202	-121	-205	-195	-258	-201	-257	-192
9	1	2	0	1	1	1	1	1	0	-190	-179	-187	-176	-180	-186	-186	-30	-241	-247	-217	-220	-222	-138	-222	-216	-213	-240	-215	-227
10	3	1	0	1	2	1	1	1	1	0	-231	-224	-200	-230	-244	-222	-65	-230	-192	-251	-211	-261	-156	-227	-248	-224	-254	-224	-253
11	3	1	2	3	2	3	2	3	1	3	0	-269	-258	-232	-203	-267	-119	-201	-174	-218	-192	-224	-146	-203	-214	-268	-224	-268	-215
12	2	1	2	3	2	3	2	3	1	2	4	0	-271	-216	-220	-285	-103	-191	-181	-211	-201	-218	-125	-215	-207	-255	-218	-256	-210
13	1	1	2	3	1	3	2	3	1	1	3	3	0	-219	-195	-266	-96	-173	-171	-190	-186	-198	-117	-198	-190	-249	-198	-253	-191
14	3	1	1	2	2	2	1	2	1	3	4	3	2	0	-203	-218	-67	-203	-177	-216	-190	-222	-155	-204	-216	-224	-219	-225	-213
15	2	1	0	1	2	1	1	1	1	2	2	2	1	2	0	-220	-43	-204	-189	-230	-209	-240	-134	-226	-229	-203	-234	-203	-230
16	2	1	2	3	2	3	2	3	1	2	4	4	3	3	2	0	-100	-189	-183	-208	-199	-215	-125	-213	-208	-255	-216	-255	-209
17	2	0	2	2	1	2	1	2	0	2	4	3	2	3	1	3	0	-63	-36	-59	-38	-60	-57	-39	-60	-110	-61	-115	-67
18	3	2	0	1	2	1	1	1	3	3	3	2	1	3	2	2	2	0	-257	-290	-252	-290	-208	-255	-279	-271	-310	-269	-300
19	1	2	0	1	1	1	1	1	3	1	1	1	1	1	1	1	0	3	0	-255	-257	-261	-178	-256	-248	-242	-264	-245	-254
20	3	1	0	1	2	1	1	1	1	3	3	2	1	3	2	2	2	3	1	0	-286	-385	-292	-322	-368	-353	-353	-350	-325
21	1	2	0	1	1	1	1	1	3	1	1	1	1	1	1	1	0	3	3	1	0	-293	-197	-286	-278	-271	-292	-272	-276
22	3	1	0	1	2	1	1	1	1	3	3	2	1	3	2	2	2	3	1	3	1	0	-298	-327	-381	-360	-365	-360	-337
23	2	0	1	1	1	1	0	1	0	2	3	2	1	3	1	2	3	2	0	3	0	3	0	-232	-292	-291	-268	-285	-239
24	0	1	1	1	0	1	0	1	2	0	1	1	1	1	0	1	1	2	2	0	2	0	1	0	-314	-306	-320	-307	-296
25	3	1	0	1	2	1	1	1	1	3	3	2	1	3	2	2	2	3	1	4	1	4	3	1	0	-353	-355	-351	-317
26	3	1	2	3	2	3	2	3	1	3	5	4	3	4	2	4	4	3	1	4	1	4	4	3	4	0	-342	-413	-308
27	3	2	0	1	2	1	1	1	3	3	3	2	1	3	2	2	2	5	3	3	3	3	2	3	3	3	0	-345	-362
28	3	1	2	3	2	3	2	3	1	3	5	4	3	4	2	4	4	3	1	4	1	4	4	1	4	6	3	0	-309
29	3	2	0	1	2	1	1	1	3	3	3	2	1	3	2	2	2	5	3	3	3	3	2	3	3	3	5	3	0

present study, the use of the algorithm for solving linear equations did not provide substantial individual deviations from the hypothetical model proposed.

For example, the high- and low-scoring examinees solved the equations in similar ways. The strength of this similarity was influenced by the differential weights of the PRs. Similarity between the $\mathbf{Q}_{low}$ and $\mathbf{Q}_{high}$ matrices would have indicated agreement in the solution of the items that involved more difficult rules (i.e., rules with high values in the ηs). A low degree of agreement would have indicated correspondence in the solution of items with rules with low ηs. Thus, the correspondence between these matrices was contingent on the relative importance of the PRs in the solution of the items as indicated by the ηs. They represented different relative weightings of the PRs involved in the solution of the items, and contributed to the degree of similarity between the pattern of responses of both groups.

This agreement between the responses of the low- and high-scoring subgroups also could have been influenced by the fact that the overall performance of these groups in the test was relatively high. The mean scores of the low- and high-scoring subgroups were 17 and 27, respectively [i.e., approximately 1 SD (5.38) below and 1 SD above the mean of 23, respectively]. Thus, the performance of the low-scoring examinees was not that poor compared to the high-scoring examinees. This might be the effect of training in the skills involved in solving the equations, the nature of the algebraic algorithm tested, or the period of time when the test was administered (i.e., close to the end of the academic year). This suggests that the QA procedures might be more useful in earlier stages of the learning process to determine if the cognitive rules are used.

Finally, the results of this study showed that the set of eight PRs proposed was a valid representation of how the examinees solve linear equations with one variable. This supports the role of the ηs as indicators of the contribution of each PR to the item difficulties, as well as the relative difficulty of each PR in the item solutions. Thus, the ηs can be employed as guidelines for constructing new test items. Having PRs that represent how the items are solved and ηs attached to each gives the test developer valuable information for constructing items in the domain of solving linear equations. For instance, the test developer may use the values of the ηs to guide the writing of items with specific difficulty levels. These new items should then have difficulty parameters based on the particular combinations of PRs required to solve them.

In addition, the connection between the LLTM and the cognitive psychology notion of PRs used for representing the procedures for solving a mathematical task provides the theoretical framework for developing achievement tests with similar approaches. Other cognitive components or operations could be defined from mathematics education theories (e.g., Carpenter, Moser, & Romberg, 1982; Van Hiele, 1986) and used to construct items, define a **W** matrix, analyze the items using the LLTM framework, and validate this structure using QA techniques.

In this research, the LLTM and QA techniques were used as complementary procedures rather than as competitors. The findings of the QA analyses illuminate one of the most critical issues of the application of the LLTM—the definition and validation of the cognitive structure of the test (the **W** matrix). The LLTM permits the researcher to specify a priori the theoretical structure of the test. This structure is taken as the "true" structure because the model does not provide a means to validate it. However, the QA paradigm gives the researcher a statistical tool for confirming the presence of this structure in the examinees' responses to the test items. The QA procedures permit the incorporation of the information about what operations the examinees used in solving the items, and comparison with the researcher's theoretical structure of the test. Consequently, the validation of the cognitive structure of the test provides evidence for supporting the ηs as a numerical description of the difficulty of the operations included in the **W** matrix. Researchers interested in designing tests that assess cognitive components to a fine degree of specificity should consider the benefits of applying both the LLTM and QA to test construction and analysis.

References

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38,* 123–140.

Andersen, E. B. (1980). *Discrete statistical models with social sciences applications.* New York: North-Holland.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369–406.

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: Freeman.

Anselin, L. (1986). *MICROQAP: A microcomputer implementation of generalized measures of spatial association* [Computer program]. Santa Barbara: Department of Geography, University of California.

Baker, F. B., & Hubert, L. J. (1977). *Data analysis: A unified approach.* Madison: Laboratory of Experimental Design, University of Wisconsin.

Brumfiel, V., Golden, N., & Heins, M. (1986). *Prealgebra: Skills, problem solving and applications.* Orlando FL: Harcourt Brace Javonich.

Carpenter, T. P., Moser, J. M., & Romberg, T. A. (Eds.) (1982). *Addition and subtraction: A cognitive perspective.* Hillsdale NJ: Erlbaum.

Dolciani, M. P., Brown, R. G., & Cole, M. (1988). *Algebra: Structure and method.* Boston: Houghton Mifflin.

Dolciani, M. P., Sorgenfrey, R. H., & Graham, J. A. (1988). *Pre-algebra: An accelerated course.* Boston: Houghton Mifflin.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–197.

Embretson, S. E., & Wetzel, D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement, 11,* 175–193.

Ericsson, A. K., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37,* 359–374.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests* [Introduction to the theory of psychological tests]. Bern: Huber.

Fischer, G. H. (1977). Linear logistic test models: Theory and application. In W. F. Kempf & H. Spada (Eds.), *Structural models of thinking and learning* (pp. 203–225). Vienna: Hans Huber.

Fischer, G. H. (1978). Probabilistic test models and their applications. *German Journal of Psychology, 2,* 298–319.

Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika, 48,* 3–26.

Fischer, G. H., & Formann, A. K. (1972). *An algorithm and a FORTRAN program for estimating the item parameters of the linear logistic test model* (Research Bulletin No. 24). Vienna: University of Vienna, Institute of Psychology.

Fischer, G. H., & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement, 6,* 397–416.

Foster, A. G., Rath, J. N., & Winters, L. J. (1990). *Algebra one.* Columbus OH: Merrill.

Gagné, E. D. (1985). *The cognitive psychology of school learning.* Boston: Little, Brown.

Green, K. E., & Smith, R. M. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics, 12,* 369–381.

Gustafsson, J-E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology, 33,* 205–233.

Hornke, L. F., & Habon, M. W. (1986). Rule-based item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement, 10,* 369–380.

Hubert, L., Golledge, R. G., & Costanzo, C. M. (1981). Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis, 13,* 224–234.

Hubert, L., Golledge, R. G., Costanzo, C. M., Gale, N., & Halperin, W. C. (1984). Nonparametric tests for directional data. In G. Bahrenberg, M. M. Fischer, & P. Nijkamp (Eds.), *Recent developments in spatial data analysis* (pp. 171–189). Brookfield VT: Gower.

Hubert, L., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical Statistics, 29,* 190–241.

Koponen, R. (1983). *An item analysis of tests in mathematics applying logistic test models.* Unpublished doctoral dissertation, University of Jyväsklä, Finland.

Lewis, C. (1981). Skill in algebra. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85–110). Hillsdale NJ: Erlbaum.

Mayer, R. E. (1980). *Different solution procedures for algebra word and equation problems* (Technical Report Series in Learning and Cognition, Report No. 80-2). Santa Barbara: University of California, Department of Psychology. (ERIC Document Reproduction Service No. ED 205 401)

Mitchell, K. J. (1983). *Cognitive processing determinants of item difficulty on the verbal subtests of the Armed Services Vocational Aptitude Battery* (Report No.

ARI-TR-598). Arlington VA: Army Research Institute for the Behavioral and Social Sciences. (ERIC Document Reproduction Service No. ED 260 122)

Mislevy, R. J., & Bock, R. D. (1989). *PC-BILOG: Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville IN: Scientific Software.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs NJ: Prentice Hall.

Price, J., Rath, J. N., & Leschensky, W. (1989). *Prealgebra: A problem solving approach.* Columbus OH: Merrill.

Resnick, L. B., & Ford, W. W. (1981). *The psychology of mathematics.* Hillsdale NJ: Erlbaum.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning, and restructuring: Three models of learning. In J. W. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition* (pp. 37–53). Hillsdale NJ: Erlbaum.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* San Diego CA: Academic Press.

Sheehan, K., & Mislevy, R. J.(1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement, 27,* 255–272.

Sleeman, D. (1984). An attempt to understand students' understanding of basic algebra. *Cognitive Science, 8,* 387–412.

Spada, H. (1977). Logistic models of learning and thought. In H. Spada & L. F. Kempf (Eds.), *Structural models of thinking and learning* (pp. 227–261). Bern: Huber.

Spada, H., & McGaw, B. (1985). The assessment of learning effects with linear logistic test models. In S. E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics* (pp. 169–194). Orlando FL: Academic Press.

Van Hiele, P. M. (1986). *Structure and insight.* Orlando FL: Academic Press.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45,* 479–494.

Whitely, S. E., & Schneider, L. M. (1981). Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement, 5,* 383–397.

Acknowledgments

This research was part of the doctoral dissertation of the author at the University of Wisconsin. The author gratefully acknowledges Frank B. Baker's advice, encouragement, and valuable suggestions during the course of this investigation.

Author's Address

Send requests for reprints or further information to María Medina-Díaz, University of Puerto Rico, Faculty of Education, P.O. Box 23187, UPR Station, San Juan PR 00931-3187, U.S.A.