

Rule-Based Item Bank Construction and Evaluation Within the Linear Logistic Framework

Lutz F. Hornke
Aachen Technical University

Michael W. Habon
Dornier GmbH
Friedrichshafen, West Germany

In cognition research, item writing rules are considered a necessary prerequisite of item banking. A set of 636 items was constructed using prespecified cognitive operations. An evaluation of test data from some 7,400 examinees revealed 446 homogeneous items. Some items had to be discarded because of printing flaws, and others because of operation complexion or other well-describable reasons. However, cognitive operations explained item difficulty parameters quite well; further cross-validation research may contribute to an item writing approach which attempts to bring psychological theory and psychometric models closer together. This will eventually free item construction from item writer idiosyncrasies.

Item bank construction is all too often thought of as an art which requires very skilled subject matter experts. However, in common item bank construction practice, items are written by a single subject matter expert, compiled in a pilot test booklet, and tried out on a sample of examinees; then item statistics determine whether an item is to be included or excluded from a future test. Seen from this perspective, test construction might be artistic but does not give careful consideration to the psychological trait represented in any item at construction time. This aspect is considered essential for useful test interpretation and application.

In contrast, this article describes one attempt to

bring together psychological trait theory, item construction, and item analysis. To achieve this, results from cognitive psychology concerning cognitive operations are used in specifying item writing rules. Development of an entire item bank is attempted by means of these rules, which systematically combine cognitive operations. It is hoped that such rules will help to construct new items whose degree of difficulty is largely estimable from the cognitive demands of the item prior to empirical item tryout. Hence a cognitive model of item difficulty is developed and tested that is expected to allow item domain construction and evaluation as well as test score interpretation in regard to qualitative and quantitative aspects of the psychological trait in question.

The research reported below pertains to item banking in that it emphasizes that a large set of items is needed for adaptive testing where sufficient items are available at relevant ability levels. Inspecting adaptive testing protocols (Hornke, 1981) revealed that items which are very difficult or very easy for an examinee contribute very little to measurement precision and to measurement economy, both relevant targets. The item bank envisioned here is a superset of homogeneous items from which appropriate items will be drawn as testing of any individual progresses. Thus, the item bank should contain many widely scattered homogeneous items.

In the past there appeared several attempts to systematize and improve item writing for the sake

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 10, No. 4, December 1986, pp. 369-380
© Copyright 1986 Applied Psychological Measurement Inc.
0146-6216/86/040369-12\$1.85

of objectivity and trait relatedness, especially in the field of education where Bloom or similar taxonomies were used as guidelines along with subject matter considerations. The aim was to design items in a rational and objective manner. Roid and Haladyna (1982) appear to be the last to address this problem comprehensively. In psychological measurement, recent work by Whitely and Schneider (1981), Embretson (1983), Fischer and Formann (1972), and Fischer and Scheiblechner (1970) attacked item construction from two directions: formal psychometric models and substantive psychological theories.

The impetus for the present paper was the notion that it ought to be possible to devise rules by means of which nearly anybody could construct a new item of the Raven matrix type or achieve this by means of a computer program. However, rules which merely addressed "how to draw" an item would have been insufficient. It was felt necessary to apply cognition research to matrix items in order to devise item writing rules that draw on substantive psychological theory. Studies by Ward and Fitzpatrick (1973), Jacobs and Vandeventer (1972), and Putz-Osterloh (1981), along with the operations-based approaches of Nährer (1980a, 1980b) and Formann (1973), represent the frame of reference for the research reported below.

Method

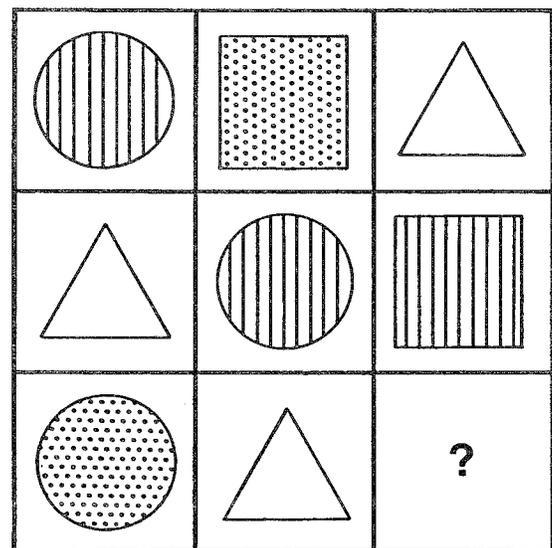
Item Construction Rules

Typical features of an item of the matrix type are circles, squares, and triangles which are shaded, hatched, or left blank, as well as organized in rows, columns, or scattered. For example, an item writer might wish to organize circles, squares, and triangles by row but "paint" them with dots, hatches, and other textural features by column. The resulting item, as shown in Figure 1, might be the result of two separate design rules, one concerning shape organized in rows and the other concerning texture organized in columns. Because shape and texture are used independently of each other, they might be conceived of as independently added item components. Keeping them apart and presenting them

separately would result in easier items than the composite item of Figure 1. This difficulty argument follows the notion that separating the two embedded items would require more mental labor because their assembly by row and by column is to be "discovered" by the examinee. However, it is assumed that any examinee can cope with shape and texture; these two aspects by themselves will not contribute greatly to item difficulty as such. The examinee's task (namely, separating item components) is nothing but the reverse of the item writer's task (adding item components in row and column directions, i.e., applying an item construction rule).

Formann (1973) designed a tripartite system based on (1) shape, texture, number, arrangement; (2) seriation, variation, superposition; and (3) orientation by row, by column, and by row and column. Accordingly, the construction rule of the item presented in Figure 1 would read: "Select shapes (three different ones from a set of shapes) and organize them in series (one scheme from a set of organizing schemata) of three rowwise (one direction from a set of directions), then select textures (dots, hatches, and blank in this case) and organize them in series

Figure 1
 Example of a Matrix-Type Item
 Stem Constructed According to Rules



of three columnwise.' It is obvious that the use of any alternative shapes and textures would result in items of hypothetically equal difficulty, because the mental operations needed to discern components separately are the same. Using squares in the first row and triangles in the second ought not to make any difference, because the construction rule stays the same as well as the mental labor of separating item components and finding their relations.

The supposedly compositional nature of the item solution led Formann (1973) to apply the linear logistic test model (LLTM; Fischer, 1974, pp. 340–406) to evaluate test data with his type of items. This approach was adopted here for reasons of comparability. The LLTM is an elaboration of the more general Rasch model (Fischer, 1974) and contains one (sub-)parameter for each aspect of the item within the tripartite system described above. What makes an item difficult is the number and the nature of the cognitive demands involved. Within the Rasch model, used in the studies referred to above, item difficulty is defined by one parameter:

$$P_{ij}(+) = [\exp(\theta_i - b_j)] / [1 + \exp(\theta_i - b_j)] \quad (1)$$

However, its decomposition in regard to the number of cognitive demands of an item yields the LLTM:

$$P_{ij}(+) = \frac{\exp\left[\theta_i - \left(c + \sum_g^k o_g d_{gj}\right)\right]}{1 + \exp\left[\theta_i - \left(c + \sum_g^k o_g d_{gj}\right)\right]} \quad (2)$$

In Equation 1 b_j represents the item's difficulty which is decomposed in Equation 2 into a sum of structural parameters, o_g , reflecting the difficulties of the hypothesized mental operations. The latter are considered necessary to solve the item. However, because the item in Figure 1 consists of four different mental operations (defined below), this item's overall difficulty, b_j , is the sum of all k mental operations difficulties, o_g , weighted by respective item design vector elements, d_{gj} . The design vector is a coding vector indicating how often an operation is included in an item. Thus, for the item bank of this study, a design matrix was constructed which represents the operation contents of

all items and thereby defines the content validity of the entire bank as such. In Equations 1 and 2 θ_i represents the ability parameter of person i , and in Equation 2 c is a norming constant.

Table 1 explains the use of the LLTM in Equation 2 for the item in Figure 1. It might be argued that the three-parameter model may be more appropriate than the Rasch model. This may be the case when the number of distractors is small and the probability of correct guessing is substantial. However, with the item format chosen here, eight alternatives are constructed and displayed; some of these represent part solutions. Thus the basic probability of guessing could be set to 1/8, a small number compared to most standardized tests.

Results from previous empirical studies showed that although items were constructed according to rules such as those described above, they did not conform entirely to the homogeneity assumptions made by the model and the expectations of the item writers. Nährer (1977), who used newly con-

Table 1
 Application of the Linear Logistic Model
 to the Test Item in Figure 1

Part	Mental Operation	Operation	Parameter
1	Identify relevant shape composition	Variation of Closed Gestalts	o_1
	Find how shapes are arranged	Columnwise	o_2
2	Identify relevant pattern composition	Variation of Closed Gestalts	o_3
	Find how patterns are arranged	Columnwise	o_2
	Find how parts are composed	Embedded Components	o_4
	No other operations included, $d_{gj...kj} = 0$	$\sum_{g=5}^k (o_g d_{gj}) = 0$	
			$b_j = c + \sum_{g=1}^k o_g d_{gj}$

structed items, could show that his items fitted the additive model to a lesser degree than expected. However, in his study the relation of nine item design features to the number of items (41) was unsuitable to test such an important item design hypothesis. Post hoc redefinition of item operation structures led to more favorable results. Although a refined structure yielded homogeneous items, fitting the LLTM did not give satisfactory results (Hornke & Habon, 1984).

Published discussions of matrix-type items (Jacobs & Vandeventer, 1972; Putz-Osterloh, 1981; Ward & Fitzpatrick, 1973) led to a revised set of rules which retained some elements from previous studies but elaborated others. The procedure was (1) to formulate item construction rules, (2) to write a complete bank of items by means of crossing all cognitive operations formulated, and (3) to evaluate item difficulty parameters in light of the hypothesized operations.

Items comprising two independent and simpler part items or components were constructed according to the following rules:

1. Eight ways to compose (part) items are defined: Identity, Addition, Subtraction, Intersection, Unique Addition, Seriation, Variation of Open Gestalts, and Variation of Closed Gestalts. (These are exemplified by the one-component items in Figure 2.)
2. Relations may occur by row, by column, or by row as well as column (see the upper three items in Figure 2 respectively, for examples).
3. Both (part) items may be separated, integrated, or embedded. Here Separation means that the two are clearly discernible. In Integration the task is to distinguish the two mentally different dimensions of the item (e.g., shape and texture). However, Embedding requires additional mental search operations. The examinee must discover which part of a gestalt might be one or the other (part) item that follows its own rule. (See the remainder of Figure 2 for examples.)

A complete set of 648 structurally different items should have resulted. However, 32 items using

Identity could not be constructed. Student item writers were introduced to the use of the above-mentioned set of rules; they were given a set of standardized symbols (e.g., different sized squares) and they constructed all 616 items possible. From their perspective, item construction consisted merely of combining prespecified cognitive operations from a list and referring freely to a set of standardized graphic features; hence, there was little chance that idiosyncrasy could enter at large. Besides the item stem, the set of eight alternatives was also based on rules. Solutions to each of the (part) items or other supposedly attractive aspects of the stem were included as distractors.

Empirical Design

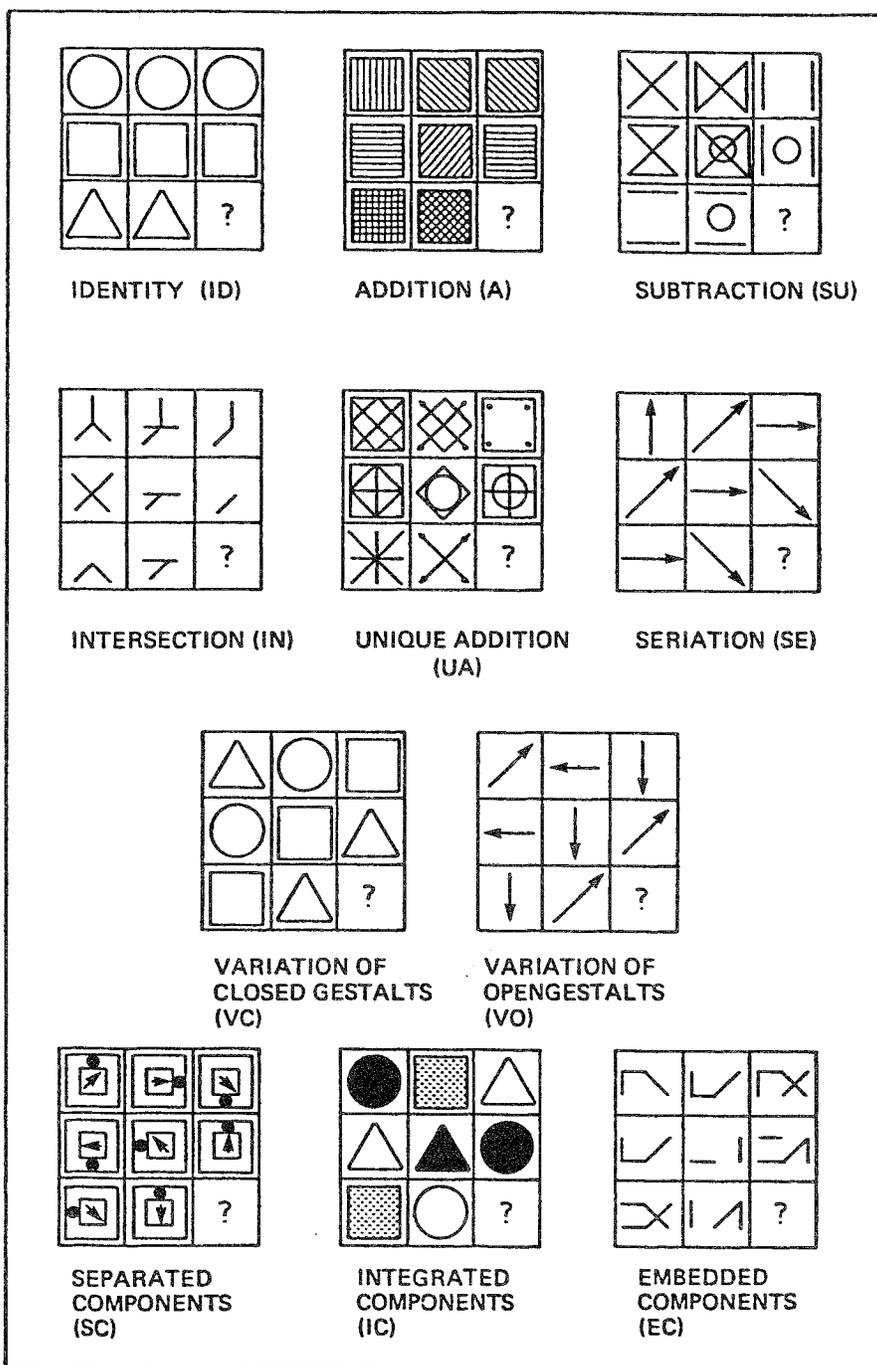
Because a set of 616 items could never be presented to any single examinee, the set was broken down into 35 smaller tests of 24 items each. In order to assure comparability of item statistics, 12 items in each test served as the link set and the other 12 as the tag set which was unique to each test. All 35 tests were organized in a chain so that each was linked by 6 items to its neighbor, according to suggestions by Wright and Stone (1979). A total of 7,400 recruits of the German Armed Forces were tested, yielding an average of 211 examinees per item.

Results

The major aim was to determine whether the use of cognitive operations in item construction yielded stable and reasonable contributions to overall Rasch item difficulty. As a first step, within each subsample of tests and examinees it was necessary to demonstrate that the homogeneity assumption of the Rasch model was met. A lack of homogeneity might have resulted from a misfit of the item itself or from examinee noncompliance with testing conditions. Thus, both sources of nonhomogeneity were studied using individual person \times item misfit scores given by Wright and Stone (1979, p. 69ff) as z^2 , defined as

$$z^2 = \exp[(2u - 1)(b - \theta)] \quad , \quad (3)$$

Figure 2
 Set of Operations Used in Item Construction



where $u = 1$ for a correct response or $u = 0$ for an incorrect response.

Table 2 displays a minor part of the overall matrix of z^2 s, demonstrating that overall item misfit might be traced back to person misfit by inspecting test data closely. The total score of person i showed that item solutions for Items 14 and 24 did not agree with his or her overall score level. However, this individual malcompliance contributed 420 and 194.3 to the item misfit statistic, respectively. Removing person i from the dataset greatly reduced the significance of the item misfit. Hence, studying person misfit, item misfit, fit using different ability groups, item fit taking subpopulations into account, and other aspects led to a multiple criterion concerning whether to remove examinees, items, or both.

From the total 616 items, 134 items (21.1%) were found to be nonhomogeneous after removing 360 response vectors (i.e., 5% of the examinees). This large number of misfitting items, however, called into question the general validity of the item construction approach. Therefore, all remaining and heretofore misfitting items were reinspected and classified according to their operational content. It was found that some graphic representations were far too complex, that is, too many graphic features were overlaid. Some items could be solved using a gestalt strategy instead of the operations intended at the time of construction. Some gestalts would

lead to problems in discerning Component 1 and Component 2 (gestalt decomposition). Some items were lacking figural features such as arrows or dots, which apparently fell off the paper before printing.

A number of problems were identified in the removed items that resulted in nonhomogeneity. These included unclear items (5.5% of the total number of items), graphics that were too complex (.8%), construction faults in items having integrated components (.8%), items that could be solved using a gestalt strategy (2.2%), problems with gestalt decomposition (6.1%), and other construction flaws (10.0%). Other items had problems in construction rules. For example, Subtraction appeared to be problematic in 4.6% of the items. Some items used the rule "subtract row 1 from row 2"; others used "subtract row 2 from row 1". Two different tasks are required; hence the subtraction rule actually comprises two rules. Items constructed by means of the latter rule were discarded from the analysis in order to remain unequivocal. Another problem (11.9% of the items) arose from items using Identity in combination with Addition, such as when both were applied in the same direction (e.g., vertically). These items are, in effect, one-componential and thus demanded less mental effort to solve. Such items were not discarded but were reclassified as one- rather than two-componential. When these problems were resolved, a set of 446 homogeneous items (70% of the original set) re-

Table 2
 Item Misfit Caused by Individual Person Misfit

Person	Item		Total
	$b_{14} = 2.44$	$b_{24} = 1.67$	
$\theta_i = -3.60$... $z_{i,14}^2 = 420$... $z_{i,24}^2 = 194.3$	$\chi_{23}^2 > 614.3^{**}$
	.	.	
	.	.	
$\theta_j = -2.92$ $z_{j,24}^2 = 98.6$	$\chi_{23}^2 > 98.6^{**}$
	.	.	
	.	.	
Total	$\chi_{200}^2 = 551.5^{**}$	$\chi_{200}^2 = 529.0^{**}$	
Removing Persons i and j	$\chi_{198}^2 = 131.5$	$\chi_{198}^2 = 236.1$	

**Statistically significant at $p < .01$

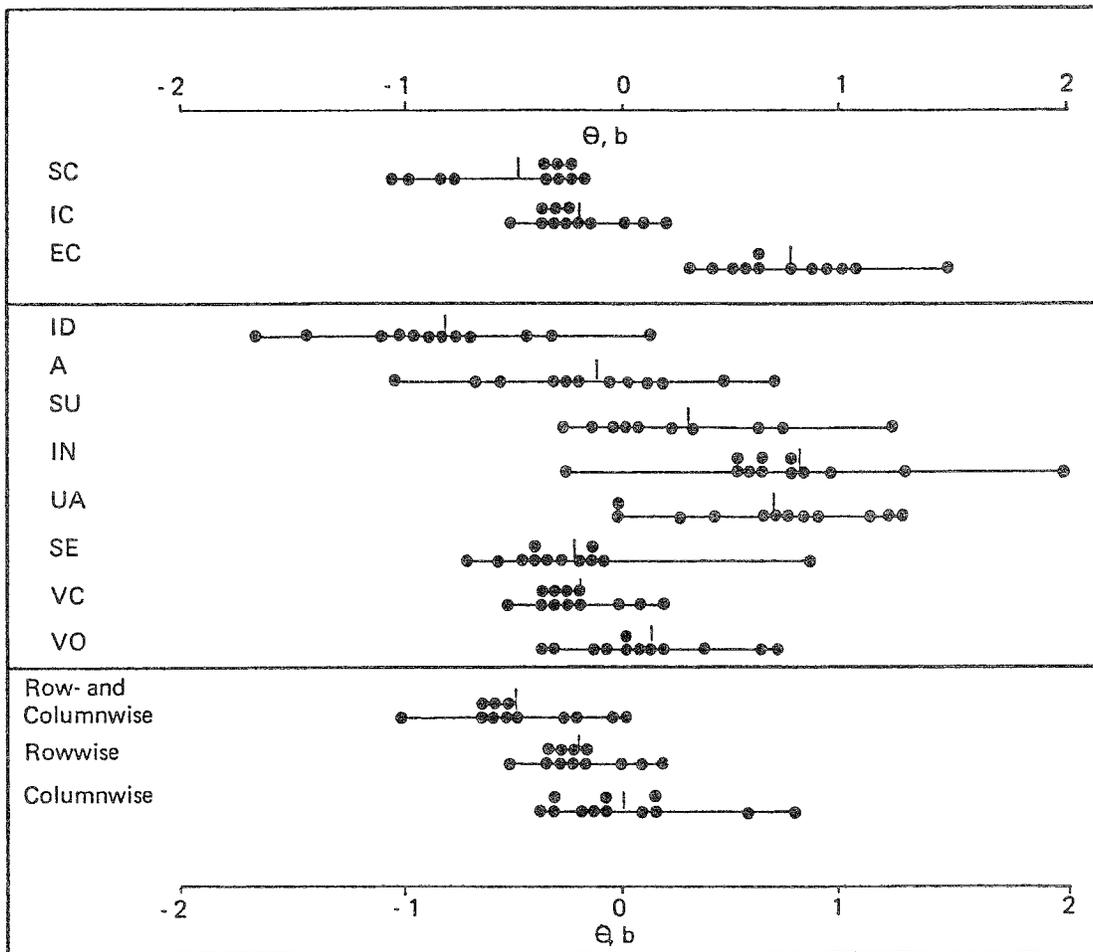
mained, which supplied data for further difficulty decomposition using operation (content) hypotheses.

The first step toward a large bank of reasonably homogeneous items was thus achieved. The second step was to estimate the parameter for each operation as a partial contribution to overall item difficulty. However, it was not possible to use either all item data or all person data at the same time. Therefore, 12 subdivisions were organized such that three linked tests were combined into one data subset, which guaranteed that all 14 operations were

included repeatedly. Applying the LLTM program by Fischer and Formann (1972) resulted in the distributions of estimates of the operation parameter values shown in Figure 3.

It is obvious that some operation parameter estimates were quite satisfactory, as can be seen from the denser clusters. However, Intersection parameter estimates varied over a range of more than 2 scale units, with a tendency to cluster around .75. This might be due partly to the number of items in the set. Because of item deletions, one set contained only 27 items where a maximum of 60 items

Figure 3
 Distribution of Operation Parameters from 12 Persons-Items Subsets
 (Vertical Lines Characterize Operation Parameter Means; For Abbreviations See Figure 2)



would have been possible. This, of course, restricted the number of times an operation appeared in the set and was combined with some other operation. In addition, at several test sites, test administration conditions were not satisfactorily maintained, and this may also have been a factor. Thus, person-item subsets larger than 60 items by 600 examinees were necessary for analysis.

The LLTM computer program could not, however, handle such large matrices or even the total dataset; multiple regression was therefore used, with Rasch difficulty estimate as a dependent variable and item design variables as independent variables. (For the abovementioned 12 data subsets, the correlation between multiple correlation-based and LLTM-based operation parameter estimates yielded a median of $r = .99$, with a minimum of $r = .76$ and a maximum of $r = .99$.) The multiple correlation for the entire dataset was $R = .65$, which means that about 40% of the variance in the Rasch parameters is attributable to the 14 cognitive operations used in the item construction rules described above.

Discussion and Conclusions

Previous studies using Raven-type items attempted to describe mental load by means of mental/cognitive operations that were considered necessary to cope with them. Within the framework of the Rasch psychometric model and its linear logistic elaborations, studies spanning 10 years yielded hints that, at least at an ordinal level, incorporating hypothetical cognitive operations does indeed contribute to item difficulty and might thus be used to guide item construction. Data from these studies—Formann, 1973; Pischwanger, 1975; Nährer, 1977; Habon, 1981; Hornke and Habon, 1983—are integrated in Figure 4.

It ought to be emphasized that all studies used newly constructed items which were based on preestablished rules. Although single operation contributions were not stable across studies, Figure 4 shows an interesting trend of operation parameter estimates, even though items were tried out at different times, at different locations, and designed

by quite independent researchers. In the present study, item writing rules were used to write a set (population) of items by means of a set of completely crossed operations. Both item stems and response alternatives were designed by means of rules. The latter represented full solutions, part solutions, and alternatives which are intended to lead the examinee astray because of highly attractive operations or features of the stem. One major achievement is seen in the large number of homogeneous items (446). They form an item bank that is able to fulfill various testing purposes. It seems that prespecified rules and a set of standardized common symbols are important contributors to homogeneity.

Although item writing was not left to idiosyncratic approaches, flaws observed in the items revealed that student item writers and test printers did not work carefully enough; 10% of the items proved to be faulty. However, the fact that these items were discovered during statistical and model fit analysis contributed to the validity of the entire approach. Their misfit demonstrates that faultily conceived rules or improperly designed patterns can be detected by the underlying psychometric model. Other misfitting items could be traced back to different solution strategies. This calls for a revised set of item writing rules, separating out the above discovered strategies or incorporating them as was done with Subtraction.

Although the major goal was achieved, a word of caution is in order concerning the linking strategy. Each of the 35 tests was checked for model fit individually. In cases where the items linking two neighboring tests were found to fit the model in both tests, all fitting items in the two tests were projected onto a common scale. However, the assumption of a common dimension will have to be cross-validated in a future study because in one case, item deletion reduced the link set to only two items.

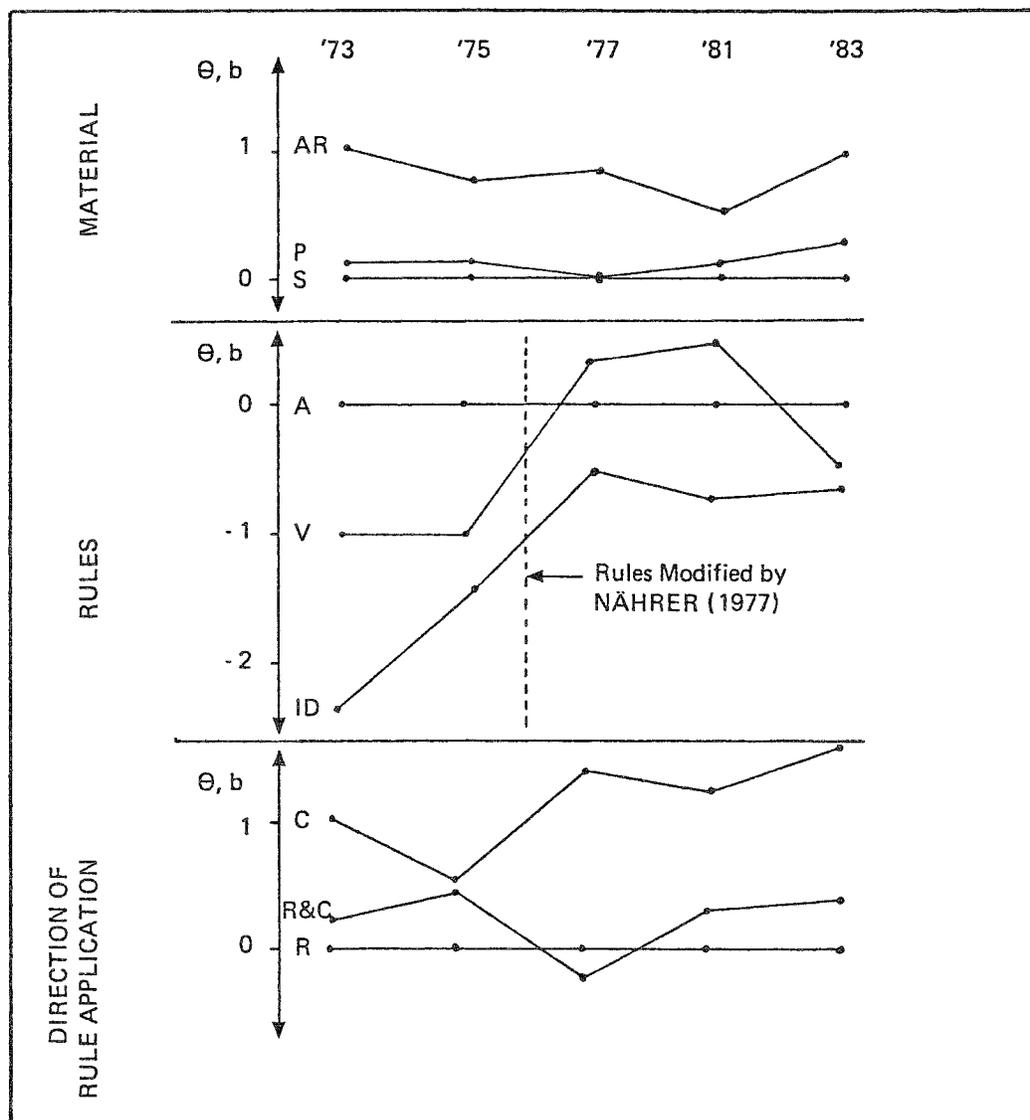
As a lower bound, 40% of the item difficulty estimates seem to be due to item writing rules (i.e., basic cognitive operations). It is an open question whether this is a large or small proportion, because no comparable studies exist which report dissimilar

results. Within this rule set, the overall difficulties of two items (displayed in Figure 5) would be estimated as $b(A) = .21$ and $b(B) = -.41$.

The difference in item difficulty estimates is said

to be due to the different directions in which Intersection and Seriation were applied: Item A "columnwise" and Item B "rowwise." Because all cognitive operations which might make a dif-

Figure 4
 Comparison of LLTM-Parameter Estimates for a Restricted Number of Operations Within a Tripartite System Commonly Used by Different Researchers
 (AR: Arrangement, P: Paint/Texture, S: Shape; A: Addition, V: Variation, ID: Identity; C: Columnwise, R&C: Row and Columnwise, R: Rowwise)



ference were known to be part of the item writing rules, this leads to an empirical test of the construction theory as such. However, whether operation interactions did contribute here could not be substantiated with the present data. It therefore seems that the linear combination of cognitive operation parameters moderates the item difficulty to a large extent. Although the linear difficulty decomposition used here was by no means perfect, other easily applied rule sets for item construction did not seem to exist. Two structurally equivalent but phenotypically different items would not be expected to have the same difficulty. However, it would be expected that they stand closer to each other on the common scale than structurally different items. As can be seen from the estimates of the operation parameters, Seriation is equal to $-.20$ and Variation of Closed Gestalts is equal to $-.23$. Neither differed very much, although quite different kinds of mental labor were involved. But this was a case where cognitive quality and degree of difficulty contribution were not substantially correlated. It may be desirable to search for additional cognitive operations that will eventually make a difference or are more highly correlated with item difficulty.

An overall picture of the results is displayed in Figure 6, which represents LLTM-based estimates of Rasch item parameter values. It can be seen that the rules presented here do not facilitate construction of "very difficult" items. Such items can be constructed by elaborating this approach with more demanding cognitions.

The present undertaking attempted to define a bank of items by means of construction rules based on substantive cognitive theory. Further development of this approach might produce items which need not be tried out empirically prior to administration. Their cognitive operation composition would suffice to moderate their difficulty and interpretability. In the area of individual personnel evaluation, the major benefit would be that items could be assembled solely according to personnel decision criteria. The achievement of this goal would require intensive basic cognition research and item analysis as well as cross-validation of the results.

References

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

Figure 5
 Calculation of Item Difficulties from Operation Difficulty Estimates

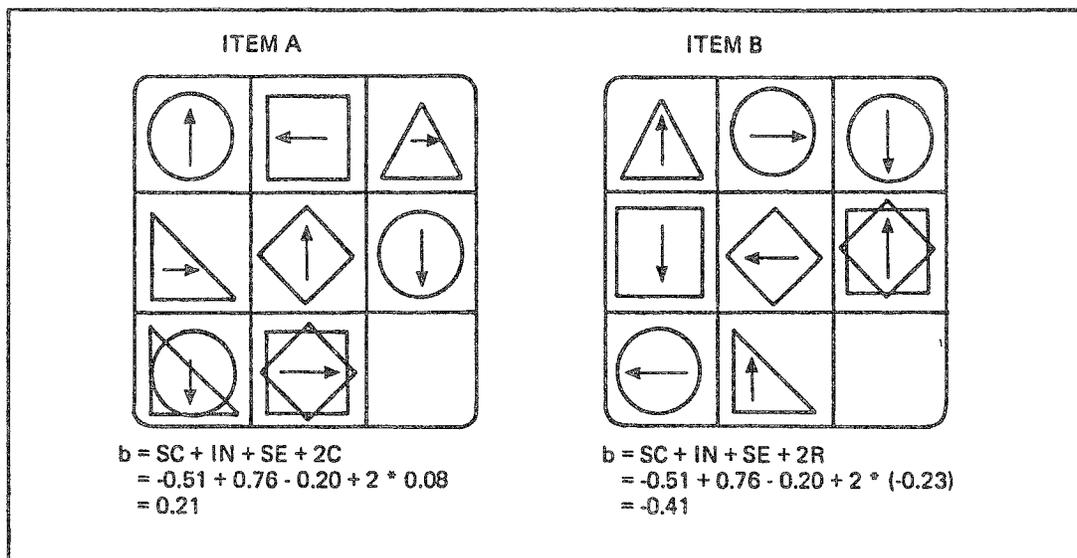
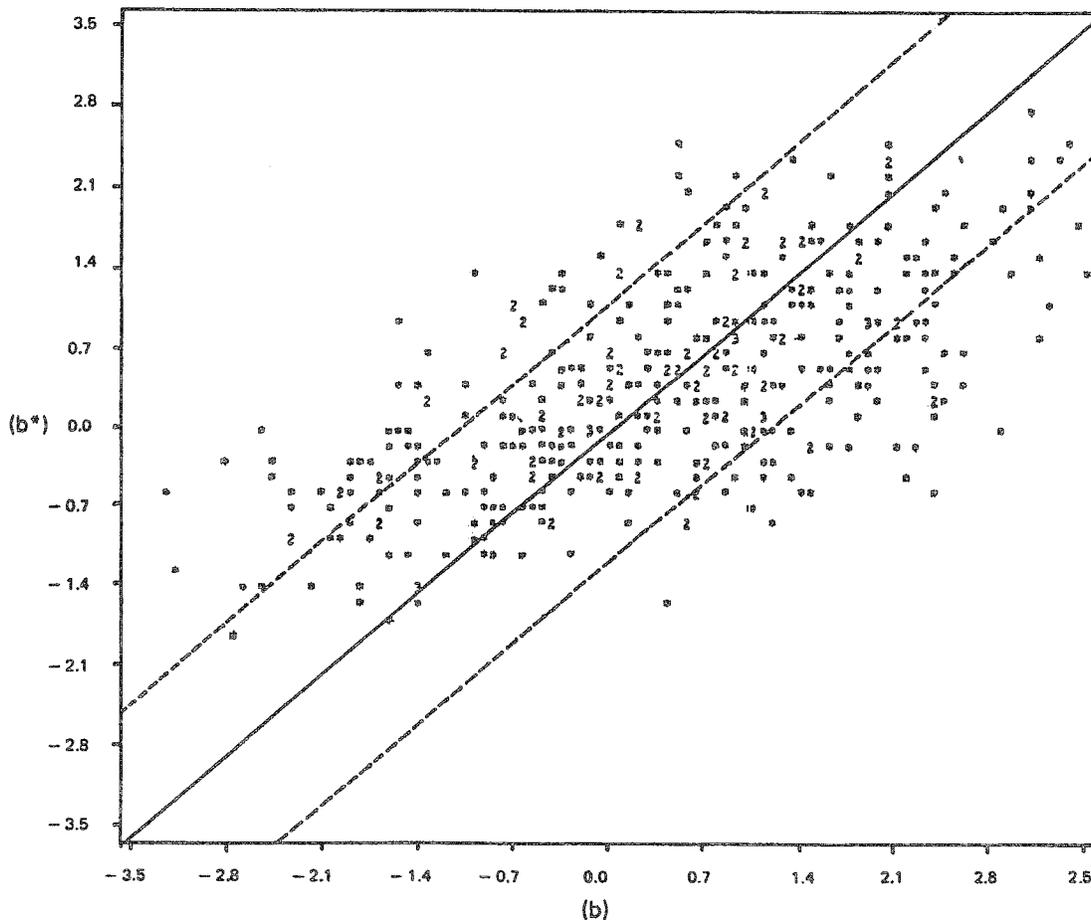


Figure 6
 LLTM-Based Item Difficulties (b^*) against Rasch Parameter Estimates (b) for 446 Items



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

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 11/72). Vienna: Psychological Institute.

Fischer, G. H., & Scheiblechner, H. H. (1970). Algorithmen und Programme für das probabilistische Testmodell von Rasch [Algorithms and programs for the Rasch probabilistic test model]. *Psychologische Beiträge*, 12, 24–51.

Formann, A. K. (1973). *Die Konstruktion eines neuen Matrizen-test und die Untersuchung des Lösungsverhaltens mit Hilfe des Linearen Logistischen Testmodells* [Construction of a new matrices test and inves-

tigation of item solving behavior by means of the linear logistic test model]. Unpublished doctoral dissertation, Universität Wien.

Habon, M. W. (1981). *Sprachfreie Tests in der pädagogischen Diagnostik: Regelgeleitete Aufgabenkonstruktion durch formale Modelle* [Language-free tests in educational diagnosis: Rule-based item construction by means of formal models]. Unpublished Diplomarbeit, Universität Marburg.

Hornke, L. F. (1981). Computerunterstütztes Testen (CUT) von Prüfungsangst [Computer-assisted testing of examination anxiety]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2, 325–335.

Hornke, L. F., & Habon, M. W. (1983). Evaluation einer rationalen Itemkonstruktion [Evaluation of rational item construction]. In R. Olechowski (Ed.), *Der*

- Beitrag der empirischen Erziehungswissenschaft zur Praxisverbesserung von Schule, Unterricht und Erziehung*. Braunschweig: Braunschweiger Studien zur Erziehungs- und Sozialwissenschaft, Band 10, Technical University of Braunschweig.
- Hornke, L. F., & Habon, M. W. (1984). Regelgeleitete Konstruktion und Evaluation von nicht-verbale Denkaufgaben [Rule-based construction and evaluation of non-verbal reasoning items]. *Wehrpsychologische Untersuchungen*, Heft 4.
- Jacobs, P. J., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement*, 32, 235–248.
- Nährer, W. (1977). *Modellkontrollen bei der Anwendung des linearen logistischen Modells in der Psychologie* [Model controls for applications of the linear logistic model in psychology]. Unpublished doctoral dissertation, Universität Wien.
- Nährer, W. (1980a). Modellkontrollen bei der Anwendung des linearen logistischen Testmodells [Model controls for applications of the linear logistic model]. *Diagnostica*, 26 (2), 112–118.
- Nährer, W. (1980b). Zur Analyse von Matrizenaufgaben mit dem linearen logistischen Testmodell [Analysis of matrix-type items with the linear logistic test model]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 27 (4), 553–564.
- Piswanger, K. (1975). *Interkulturelle Vergleiche mit dem Matrizen-test von Formann* [Intercultural comparisons using Formann's matrices test]. Unpublished doctoral dissertation, Universität Wien.
- Putz-Osterloh, W. (1981). *Problemlösungsprozesse und Intelligenztestleistung* [Problem-solving processes and intelligence test performance]. Bern: Huber.
- Roid, G. H., & Haladyna, T. M. (1982). *A technology for test item writing*. New York: Academic Press.
- Ward, J., & Fitzpatrick, F. (1973). Characteristics of matrices items. *Perceptual and Motor Skills*, 36, 987–993.
- Whitely, S. E., & Schneider, L. M. (1981). Information structure for geometric analogies: A test theory approach. *Applied Psychological Measurement*, 5, 383–397.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.

Author's Address

Send requests for reprints or further information to Lutz F. Hornke, Institut für Psychologie, Aachen Technical University, Jägerstrasse 17/19, D-5100 Aachen, West Germany.