DOCUMENT RESUME

ED 422 383	TM 028 945
AUTHOR	Yepes-Baraya, Mario; Tatsuoka, Kikumi; Allen, Nancy L.; O'Sullivan, Christine; Liang, Jo-Lin; Hui, Xuefei
TITLE	Application of the Rule-Space Methodology to the 1996 NAEP Science Assessment: Grade 4 Preliminary Results.
INSTITUTION	Educational Testing Service, Princeton, NJ.; National Assessment of Educational Progress, Princeton, NJ.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	1998-04-00
NOTE	40p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Diego, CA, April 14-16, 1998).
CONTRACT	R999J50001
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Coding; *Elementary School Students; Grade 4; Intermediate Grades; *Knowledge Level; *Research Methodology; *Science Education; *Science Tests; Tables (Data); *Test Items
IDENTIFIERS	*Rule Space Model; *State Science Assessment (NAEP)

ABSTRACT

In the context of Phase Four of the National Assessment of Educational Progress (NAEP) Science Attribute Study, this report includes a discussion of item attributes, an overview of the item attributes used in the study, some psychometric characteristics of the blocks analyzed, a general description of the rule-space methodology, the results obtained, and a discussion. Two grade-four booklets from the 1996 NAEP Science Assessment were coded in preparation for the rule-space analysis, using a total of 4 different blocks of items and 328 examinees. The application of the rule-space methodology to an assessment involves essentially two stages: (1) the identification and coding of item attributes for the items, as performed, and the determination of knowledge states; and (2) the classification of examinees into one of the predetermined knowledge states. By design, the NAEP science assessment is a balanced assessment in which all examinees are required to answer one block of items of each type (conceptual/problem solving, theme, and performance task). On the surface, this property would appear to warrant generalizability of the findings of this study about item attributes to the remaining Grade 4 blocks, but the science content does vary across blocks, and the interaction between content and other item attributes could have unforeseen results. It is suggested that the analyses performed in this study be extended to other booklets. (Contains 7 tables, 4 figures, and 14 references.) (SLD)

Application of the Rule-Space Methodology to the 1996 NAEP Science Assessment: Grade 4 Preliminary Results

Mario Yepes-Baraya
Kikumi Tatsuoka
Nancy L. Allen
Christine O'Sullivan
Jo-Lin Liang
Xuefei Hui

Educational Testing Service
Princeton, NJ 08541

Paper presented as part of the Symposium *Evaluating the 1996 NAEP Science Assessment: Major Findings and Future Directions* at the annual meeting of the National Council on Measurement in Education, San Diego, CA, April 1998.

The work reported herein was partially supported under the National Assessment of Educational Progress (Grant No. R999J50001) as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

Address correspondence to: Mario Yepes-Baraya, Ph.D., ETS (8-R), Princeton, NJ 08541. Telephone: (609) 734-5357. E-mail: myepes@ets.org

Application of the Rule-space Methodology to the 1996 NAEP Science Assessment: Grade 4 Preliminary Results

Mario Yepes-Baraya, Kikumi Tatsuoka, Nancy L. Allen, Christine O'Sullivan, Jo-Lin Liang, Xuefei Hui

Educational Testing Service, Princeton, NJ 08550

Introduction

The study described in this report corresponds to Phase Four of the NAEP Science Attribute Study, an ongoing research effort at Educational Testing Service to better understand the 1996 NAEP science assessment results. The NAEP Science Attribute Study was begun in 1992 with items from the 1993 NAEP science field test. To date, it has entailed four separate, though related, phases. Results of the first three phases led to the identification and refinement of item attributes related to item performance on the NAEP science assessment (Park, & Allen, 1994; Yepes-Baraya, & Allen, 1994; Allen, Park, Liang, & Thayer, 1995; Yepes-Baraya, 1995, 1996, 1997). These attributes (see Table 1) were used in Phase Four to code two Grade 4 booklets from the 1996 NAEP science assessment. The rule-space methodology (Tatsuoka, 1983; Tatsuoka, & Tatsuoka, 1989) was then used to identify knowledge states for a sample of 328 examinees.

This report includes a discussion of item attributes, an overview of the item attributes used in the study, some psychometric characteristics of the blocks analyzed, a general description of the rule-space methodology, the results obtained, and discussion.

What are Item Attributes?

An approximation of the notion of item attributes in science (and other subjects) can be gained by making reference to the science curriculum frameworks and standards available from each State Education Department. One reason for the development of these frameworks was to facilitate the measurement of student achievement by identifying relevant science knowledge, skills, and practices. The State of New Jersey, for example, has identified twelve general science standards: five process standards and seven content standards (New Jersey Science Curriculum Framework, September, 1997). For each standard, a number of cumulative progress indicators have been identified. As illustrated in Table 2, the language used for the progress indicators is more specific (with respect to learning and assessment) than that used for the standards, but still fairly general and not very useful for assessment purposes. Analysis of the sample assessment activity, however, can provide useful information with respect to the types of knowledge, skills, and features of the assessment activity that presumably are relevant for performance on this activity. Some of the knowledge, skills, and features of the assessment activity that may be relevant to performance on the activity are included in Table 2 under possible item attributes. Thus a working definition of item attribute is any feature of an item that may be associated with student performance on said item.

The list of possible item attributes presented in Table 2 suggests that the number and type of relevant attributes can vary depending on the type of assessment and reasons for identifying item attributes. In the example given, the attributes make reference to broad categories in the framework (e.g., physical science), general principles (e.g., laws of motion), general physical science concepts (e.g., speed, time, distance, mass), response format (e.g., constructed response), use of figural information (e.g., drawing is provided, drawing is required), use of calculations, use of explanations, whether similar problems were done in class, and whether certain experiences outside the classroom might be helpful to answer the item.

Attribute Coding of the NAEP Science Blocks

There were three sources of attributes identified for the NAEP science assessment: 1) the NAEP science framework (O'Sullivan, 1995); 2) item attributes related to item performance identified by a group of ETS researchers (Yepes-Baraya, & Allen, 1994), and 3) a cognitive model of problem-solving (Sugrue, Fall 1995). Each item in the assessment is characterized by the presence or absence of 38 attributes. An abbreviated description of the attributes is provided in Table 1. The attributes have been classified into six categories: 1) content knowledge, 2) reasoning with content and explaining, 3) hypothesis formulation and testing, 4) processing figural information, 5) item format and reading difficulty, 6) and process skills for hands-on tasks. Content knowledge pertains to items for which certain types of knowledge (e.g., knowledge of facts or concepts, or knowledge derived from practical experience) can be used or are required to answer correctly. Reasoning with content and explaining refers to items requiring some form of deductive or inductive reasoning involving science content. Items in the third category require the formulation or testing of a hypothesis. Processing figural information describes items requiring the processing of information contained in a table, graph or figure, or the provision of a figural response. Item format and reading difficulty groups items with sentence structures and format characteristics that might facilitate or hinder answering the item correctly. Process skills for hands-on tasks refers to items requiring manipulation of equipment or materials, making observations or measurements, and other science process skills.

Two Grade 4 booklets from the 1996 NAEP Science Assessment were coded in preparation for the rule-space analysis. Each booklet has three cognitive blocks of items: one conceptual/problem-solving block, one theme block, and one hands-on task. Since the same theme block and hands-on task appeared in both booklets, a total of four different blocks were used in the analysis. All items in the four blocks in the study were coded using the attributes in the six groups described above. The coding of these blocks is presented in Table 3 and summarized below:

- The conceptual/problem-solving blocks were highest in reasoning attributes, relatively high on knowledge attributes, and had no process skills attributes. One of the blocks was relatively high on figural attributes while the other one was lowest.
- The theme block was highest on knowledge and figural attributes, in the middle in reasoning attributes, and had no process skills attributes.

- The hands-on task was highest in process skills attributes, relatively high in figural attributes, and relatively low in knowledge and reasoning attributes.
- Only one block, a conceptual problem-solving block, had hypothesis testing attributes.
- The reading load was heaviest for the task items.
- A total of 367 attributes were coded for the four Grade 4 blocks. The breakdown of these attributes is as follows :

Knowledge	119	(32%)
Reasoning	43	(12%)
Hypothesis testing	3	(1%)
Figural information	111	(30%)
Reading load	79	(21%)
Process skills	12	(4%)

Psychometric Characteristics of the Blocks Analyzed

Table 4 summarizes some characteristics of the blocks analyzed. The p-values presented in the figure are based on NAEP scaled scores and were obtained with nationally representative samples ranging from about 1200 to 1600, depending on the block type, however the sample size for the application of the rule-space methodology was about 140 for each of the conceptual/ problem-solving blocks and about 320 each for the theme block and the hands-on task. While the conceptual/problem-solving blocks have a balanced number of multiple choice items and constructed-response items, the theme block and the hands-on task have a preponderance of constructed-response items. Two measures of block difficulty included in the table are the mean p-values and the mean percent of examinees not reaching items. The percent of examinees not reaching the last item is also included. All four blocks were fairly difficult, with the conceptual/ problem-solving blocks being relatively easier (mean p-value of about 49), the theme block in the middle (mean p-value of 47), and the hands-on task the hardest (mean p-value of 44). The same pattern emerges when the blocks are compared in terms of the mean percent of examinees not reaching items. The percentages range from about 3 for the conceptual/problem-solving blocks, to

6.8 for the theme block, to 15 for the hands-on task. The percentages not reaching the last item range from about 15 for the conceptual/problem-solving blocks, to 34 for the theme block, to 47 for the task. Since subjects not reaching a given item are excluded from computation of item statistics, the p-values presented in the table would be considerably lower if all subjects had been included.

The Rule-Space Methodology

As seen above, the coding of items in terms of the attributes identified provides descriptive information for each item, each block of items, and the assessment as a whole. This information can be used by researchers and practitioners interested in understanding the types of knowledge, skills, and information processing required by the 1996 NAEP science assessment.

A completed coding sheet for a block of items is called an incidence matrix (of items by attributes). The incidence matrix is the basis for the application of the rule-space methodology (Tatsuoka, 1983; Tatsuoka, & Tatsuoka, 1989). The rule-space methodology is a probabilistic approach to identifying patterns of examinee responses which can be used in conjunction with Item Response Theory to identify attributes that an examinee or groups of examinees have mastered at a specified probability level. The information thus obtained can be used for test development, diagnostic instruction, and assessment purposes (Tatsuoka, 1990).

The application of the rule-space methodology to an assessment involves essentially two stages: 1) the identification and coding of item attributes for the items in the assessment, as discussed above, and determination of knowledge states, and 2) the classification of examinees into one of the predetermined knowledge states.

Once the incidence matrix has been created, all the possible knowledge states can be determined. The rule-space methodology classifies examinees' response patterns and computes individual attribute mastery probabilities. It is assumed that when an item is answered correctly all the attributes characterizing said item have been applied correctly (Tatsuoka, 1997).

Results

Figure 1 represents the results of the application of the rule-space methodology to the four NAEP science Grade 4 blocks selected for this study. The y-axis represents *theta* (the total score), averaged by group. The range for *theta* is –3 to +3. The x-axis represents *zeta* (the degree of unusualness of an item score pattern), with positive values of *zeta* corresponding to unexpected scores. Unusual item score patterns occur when examinees answer incorrectly easy items and answer correctly harder items.

Three main progress paths were identified for the sample of 328 examinees. These three paths include all groups of 3 or more examinees. The total number of examinees classified by Paths 1-3 was 220, equivalent to 67% of the sample in the study. The remaining 108 examinees were either classified in groups of 2 or represented singular cases.

Path 1, the main path, links eight data points (1-8). Each point represents a knowledge state, as defined in the previous section. Path 2, to the right of Path 1, links four data points (1a, 3b, 3d, and 3e). These points represent intermediate knowledge states lying between points 1 and 4 on Path 1. Path 3, to the right of Path 2, links four data points (3a, 3c, 4a, and 7a). These points represent knowledge states between points 3 and 8 on Path 1.

Tables 5-7 provide a description of the knowledge states on each progress path. Each knowledge state is defined in terms of the following information:

- the mean score (*theta* averaged by group) for the knowledge state
- the number of examinees achieving that knowledge state
- the actual item attributes mastered by the examinees in that knowledge state
- nomenclature to represent mastery levels for each of the six attribute categories. Example for Path 1:

K = Content Knowledge – 2 mastery levels: K1, K2
R = Reasoning – 4 mastery levels: R1, R2, R3, R4

H = Hypothesis formulation and testing – 1 mastery level: H1
F = Figural information – 1 mastery level: F1
L = Item format and reading load – 2 mastery levels: L1, L2
P = Process skills for hands-on tasks – 1 mastery level: P1

The information presented in Tables 5-7 is presented as a diagram in Figures 2-4 and is explained in detail for each progress path below.

Progress Path 1

Path 1 (see Figure 1, Table 5, and Figure 2) links 8 knowledge states and includes 94 examinees (29% of the total sample). The lowest state (State 1) has a mean *theta* of –2.87 while the highest state has a mean *theta* of +2.20 (the range is –3 to +3). The mean *zeta* for States 1-8 on Path 1 is –0.54.

State 1 is defined by the mastery of basic content knowledge (K1), basic figural processing (F1) and basic reading load (L1). Each of these levels includes the attributes listed in Table 5 and described in Table 1. State 2 involves the mastery of one additional content knowledge level (K2). K2, in turn, depends on the correct application of Attribute 4 (Can knowledge of principles be used to answer the item?) The mean *theta* difference between State 1 and State 2 is 0.89 (quite large and equivalent to almost one standard deviation). State 3 includes the mastery of basic reasoning (R1) and one more figural information processing level (F2). R1 is defined by Attribute 8 (Can tracing cause-effect from one component to another in a system be used to answer the item?). F2 includes Attributes 23 and 27, both of which refer to student-generated information contained in a table, graph, or figure. The mean *theta* difference between State 2 and State 3 is 0.97 (again quite large and almost equal to one standard deviation). There are a total of 25 examinees in States 1-3.

State 4 is characterized by the achievement of two new levels, R2 and L2. R2 does not build on R1; it does not involve the mastery of R1 attributes. R2 is defined by mastery of Attribute 7 (Can reasoning from a general concept, principle, law be used?). L2 involves Attributes 30 and

31, both of which impose a heavier reading load on examinees. F2 at this state does not include Attribute 27. The mean *theta* for this group of 32 examinees (+0.17) is 1.18 higher than that of those in State 3. State 4's mean *theta* is only slightly higher than the mean *theta* for the group of 220 examinees classified by the application of the rule-space methodology.

State 5 is defined by the mastery of basic processing skills associated with hands-on tasks (P1). This state's mean score is 0.49 higher than the previous state's. State 6 is defined by the mastery of a new level of reasoning (R3) and the non-mastery of P1. The mean total score difference between State 6 and State 5 is 0.51. State 7 is defined by the mastery of the same attributes as State 5 plus Attribute 11 (Can thinking with models or analogies be used to answer the item?) and Attribute 27 (Does response require a graph or figure to be drawn or completed?). These two attributes appear to explain the 0.90 difference between the two states' mean *thetas*. State 8 is defined by the mastery of an additional reasoning level (R4) and the only hypothesis formulation level (H1) present in the four blocks studied. A total of 37 examinees are classified in States 5-8.

Progress Path 2

Path 2 (see Figure 1, Table 6, and Figure 3) indudes four states whose mean *thetas* lie between States 1 and 4 on Path 1. The intermediate states linked by Path 2 are 1a, 3b, 3d, and 3e. These four states classify a total of 90 examinees (27.4% of the total sample). The mean score for each of these states is negative. The mean *zeta* for the Path 2 states is –0.39.

The Path 2 states are different from their equivalent states on Path 1 (States 2 and 3) in two important respects: mastery of several reasoning attributes, including Attribute 12 (Does the item require that a response be given and the response be justified?), and non-mastery of attributes involving the processing of figural information (F-attributes). Since some of the content knowledge, reasoning, and reading load mastery levels for Path 2 were different from those on

Path 1, different nomenclature was used (e.g., K1A, R1A, L1A) to differentiate them from the Path 1 mastery levels.

Progress Path 3

Path 3 (see Figure 1, Table 7, and Figure 4) includes four states whose mean *thetas* lie between States 3 and 8 on Path 1. The intermediate states linked by Path 3 are 3a, 3c, 4a, and 7a. These four states classify a total of 36 examinees (11% of the total sample).

The lower two states have negative mean scores and the upper two positive mean scores. The mean *zeta* for the Path 3 states is –0.09, with the upper two states (4a and 7a) having positive *zetas*.

The Path 3 states are similar to the Path 1 states in that the same nomenclature for the mastery levels of attribute categories (e.g., K1, K2, R1, R2, etc.) can be used to define the states on each path. The Path 3 states, however, are characterized by the earlier mastery of process skills (P-attributes).

Trends Across Paths

Content knowledge (K), figural information processing (F), and reading difficulty (L) were the cognitive dimensions first represented on the progress paths. Mastery of the basic levels of these dimensions (K1, F1, and L1) describes the group of examinees near the bottom of the scale. These examinees are able to tackle items involving knowledge of facts, basic experimental procedures, concepts, and science vocabulary as well as basic science information that might have been gained through practical experience outside the classroom. These examinees are also able to answer simple constructed-response items and items requiring the processing of figural information but not the production of such information. These examinees are not able to answer items requiring reasoning with science content (R), formulating or testing hypotheses (H), or process skills for hands-on tasks (P).

As mean group scores increase on Path 1, examinees are able to master more advanced content knowledge and figural processing attributes (K2 and F2) as well as basic reasoning attributes (R1). However on Path 2 different kinds of knowledge and reasoning attributes are mastered (K1A and R1A). Moreover, F-attributes are not mastered on Path 2 until Stage 6 is reached. It would be interesting to find out why examinees on the intermediate stages of Path 2 can do some reasoning tasks (describing procedures to solve a problem, justifying a response, and reasoning from general concepts or principles to specific conclusions), but cannot do items involving basic figural information. One hypothesis is that these examinees, all of whom are in groups with negative mean total scores, come from learning and assessment environments where figural information processing is not emphasized but reasoning and explaining are.

As one moves up through the intermediate states, no new K, F, and L mastery levels are attained. Increments in total score are accounted for by the mastery of new levels of reasoning skills (R), laboratory process skills (P), and hypothesis formulation and testing skills (H). It is interesting to observe, for example, that P1 is not mastered on Path 1 until Stage 5 is reached and H1 until Stage 8. This may be explained by the relatively small number of Grade 4 students working on actual scientific investigations in school (O'Sullivan, & Pearlmutter, 1996). Only those examinees that have had opportunities to learn these skills can be expected to master them.

Discussion and Conclusions

The application of the rule-space methodology to the NAEP science assessment was done as part of the NAEP Science Attribute Study. As such, an evaluation of the application has to be done in the context of an evaluation of the earlier phases of the NAEP Science Attribute Study. Although a formal evaluation of the attribute study is beyond the scope of this report, some of the earlier results are discussed below to help the reader interpret the results of the present study.

The validity, reliability, and relevance of attribute-based research rest squarely on the type and nature of the attributes previously identified. In the case of the present study, considerable effort was put into surveying a wide range of sources of item attributes (Yepes-Baraya, & Allen, 1994), performing a variety of exploratory statistical analyses to understand the relationships between the science framework variables and the item attributes (Park, & Allen, 1994; Allen, Park, Liang, & Thayer; 1995), and conducting two separate validation studies. The first study involved protocol analysis of Grade 8 students performing a think aloud (Yepes-Baraya, 1996). The second study entailed teachers coding the entire assessment with the attributes previously identified (Yepes-Baraya, 1997).

The resulting attributes can be described as encompassing important science dimensions, widely accepted by science educators, cognitive scientists, and assessment experts. Additionally, most of the skills identified are general enough to be useful in testing situations beyond NAEP, e.g., large-scale tests involving the assessment of individual examinees, or classroom-based assessments . Notwithstanding these considerations, further research should be conducted on the types of attributes identified and the reliable coding of these attributes before using them on other assessments. The attributes in the figural response category (F-dimension), for example, were expressly developed for the NAEP science assessment. These attributes reflect the profusion of information presented in tables, charts, and pictures throughout the assessment , as well as the requirement that examinees complete charts, draw objects, or answer a given item by using figural information generated by the examinees themselves in previous items. Other assessments may not rely to the same extent on the F-dimension. Similar considerations apply to all the other dimensions.

By design, the NAEP science assessment is a balanced assessment; all examinees are required to answer one block of items of each type (conceptual/problem-solving, theme, and performance task). On the surface, this property would appear to warrant generalizibility of the findings of this study to the remaining Grade 4 blocks. The science content, however, does vary across blocks, and the interaction between content and other item attributes could prove to have unforeseen results. Thus it is suggested that the analysis performed in this study be extended to other booklets. Additionally, it is suggested that similar analyses be performed with Grade 8 and Grade 12 booklets in order to better understand the evolution of mastery of the science dimensions identified in the NAEP Science Attribute Study.

References

Allen, N.L., Park, C., Liang, J., & Thayer, D. (1995). Relationships between test specifications, task demands, and item attributes in a large-scale science assessment. Paper presented as part of the symposium Large Scale Science Performance Assessment and Results: Informing Test and Score Development at the annual meeting of AERA, San Francisco.

NAEP Science Consensus Project. (1993). Science framework for the 1994 National Assessment of Educational Progress. National Assessment Governing Board. Washington, DC: U.S. Department of Education.

New Jersey Department of Education (1997, September). New Jersey Science Curriculum Framework.

O'Sullivan, C., & Pearlmutter, A. (1996). The National Assessment of Educational Progress science work study. Paper presented as part of the symposium School-Based Special Studies of a Large-Scale Performance Assessment at the annual meeting of NCME, New York.

Park, C. & Allen, N.L. (1994). Relationships between test specifications, item responses, task demands, and item attributes in a large-scale science assessment. Paper presented at the annual meeting of AERA, New Orleans.

Sugrue, B. (Fall 1995). A theory-based framework for assessing domain specific problem-solving ability. Educational Measurement: Issues and Practices, 3, 29-36.

Tatsuoka, K.K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20 (4).

Tatsuoka, K.K. (1990). Toward an integration of item response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold & M.C. Shafter (Eds.) Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, NJ: Lawrence Erlbaum.

Tatsuoka, K.K. (1997, September). Rule-Space methodology. Princeton, NJ: Educational Testing Service.

Tatsuoka, M.M., & Tatsuoka, K.K. (1989). Rule-space. In Kotz & Johnson (Eds.) Encyclopedia of statistical sciences. New York: Wiley.

Yepes-Baraya, M. & Allen, N.L. (1994). The process of identifying item attributes related to item performance for the 1993 National Assessment of Educational Progress (NAEP) science field test. Princeton, NJ: Educational Testing Service.

Yepes-Baraya, M. (1995). Task analysis of science performance tasks and items: Identifying relevant attributes. Paper presented as part of the symposium Large Scale Science Performance Assessment and Results: Informing Test and Score Development at the annual meetings of NCME, San Francisco.

Yepes-Baraya, M. (1996). A cognitive study based on the National Assessment of Educational Progress (NAEP) science assessment. Paper presented as part of the symposium School-Based Special Studies of a Large-Scale Performance Assessment at the annual meeting of NCME, New York.

Yepes-Baraya, M. (1997). Lessons learned from the coding of item attributes for the 1996 National Assessment of Educational Progress (NAEP) science assessment. Paper presented as part of the symposium What Does the NAEP Science Assessment Measure?: Results from the 1996 Science Attribute Study at the annual meeting of NCME, Chicago.

Table 1. Item Attributes

Content knowledge
1. Can knowledge of facts be used to answer the item?
2. Can knowledge of experimental procedures be used to answer the item?
3. Can knowledge of concepts be used to answer the item?
4. Can knowledge of principles be used to answer the item?
5. Does item have science vocabulary that must be understood to answer item?
6. Could the info. required to answer item have been gained through practical experience?

Reasoning and explaining
7. Can reasoning from general concept/principle/law to specific conclusion be used?
8. Can tracing cause-effect from one component to another in system be used to answer item?
9. Can formal inductive reasoning be used to answer item?
10. Does item require identifying or describing a procedure to solve a problem?
11. Can thinking with models/analogies be used to answer item?
12. Does item require that a response be given and the response be justified?

Hypothesis formulation and testing
13. Is generation of hypothesis necessary to answer item?
14. Does item require ident. of variables/controls in design of test for hypothesis?
15. Does item require generating operationalized procedures for testing a hypothesis?
16. Does item require use of multiple control groups in design of test for hypothesis?

Processing figural information
17. Does item have a TGF* already completed/needs to be completed?
18. Does item refer directly or indirectly to info. in a completed & separate TGF (g/s)?
19. Does item refer to info. in a tTGF* (s)* separate from stem?
20. Does item have (or refers to info. in) a completed TGF (g/s)*?
21. When present, is it possible to use info. in completed TGF (g/s) to answer item?
22. Is it necessary to use info. in completed TGF (g/s) to answer item?
23. Is some of the info. needed to answer item in TGF (s)?
24. Is all info. needed to answer item in tTGF in block with item? [All info. is (g)]
25. Is all info. needed to answer item in tTGF in block with item? [Some info. is (s)]
26. Does response require a TGF to be drawn or completed?
27. Does response require a GF to be drawn or completed?

Item format and reading difficulty
28. Is item a 5 or 4-category constructed-response item?
29. Is item a 3 or 2-category constructed-response item?
30. Does item stem have at least 1/2/3 intratext referentials (e.g., it, this, these)?
31. Does item stem have at least 1/2/3 clauses with fronted structures?
32. Must response meet all conditions specified in stem?
33. Does item have exceptions/negations that make item complex?
34. Can item be solved by choosing the odd option out?

Process skills for hands-on tasks
35. Does item require the manipulation of equipment/materials?
36. Does item require the recording of data (observations or measurements)?
37. Does item require interpreting data collected or making inferences from this data?
38. Does item require performing numerical calculations with data collected?

*TGF = table, graph, or figure (g) = given
tTGF = text, table, graph, or figure (s) = student-generated

Table 2. From Standards to Item Attributes
(Adapted from New Jersey Science Curriculum Framework, September 1997)

Standard (#9)	All students will gain an understanding of natural laws as they apply to motion, forces and energy transformations
Progress Indicator (#1)	Demonstrate that the motion of an object can vary in speed and direction
Sample Learning Activity (Grade 2)	Moving Objects: Students are asked to predict the movement of two identical sheets of paper dropped from the same height, one kept flat and the other one crumpled into a ball. Which piece of paper will fall faster? Students record their predictions and explanations before actually conducting the experiment.
Sample Assessment Activity	A problem involving two sky divers, one with a small and one with a big parachute, jumping off from the same height at the same time (graphic is provided). Students are asked to draw the trajectory of each sky diver and explain their differences in speed when they hit the ground. (More or less specific information can be provided and/or requested).
Scoring Rubric	The development of a scoring rubric is needed to evaluate student performance.
Possible Item Attributes for Sample Assessment Activity	Physical science Laws of motion Falling objects Speed Time Distance Mass Constructed response item Drawing is provided Drawing is required Numerical calculations may be used Explanation is required Similar problem was done in class Outside experience may be useful Other attributes

Table 3. Attribute Coding for the Grade 4 Blocks Used in the Study Attribute Means - Correlations with Total Scores

	Block Types				TOTALS	Means (N = 328)	Corr. with Total Score
	C/PS 1	C/PS 2	Theme	Task			
Total No. Items	11	11	8	7	37		
ATTRIBUTES	No. Attributes by Block Type						
Knowledge							
1	7	10	8	2	27	0.97	0.61
2	4	-	-	5	9	0.56	0.21
3	9	9	8	4	31	0.60	0.40
4	3	7	7	4	21	0.92	0.19
5	3	6	7	-	16	0.91	0.37
6	3	4	6	3	16	0.95	0.61
Subtotal	29	36	36	18			
Reasoning							
7	3	7	8	4	22	0.60	0.51
8	2	1	-	-	3	0.23	0.15
9	-	-	-	-	-	-	-
10	3	1	-	-	4	0.37	0.09
11	3	1	-	-	4	0.21	0.43
12	4	1	2	3	10	0.43	0.13
Subtotal	15	11	10	7			
Hypothesis Testing							
13	-	-	-	-	-	-	-
14	3	-	-	-	3	0.08	0.31
15	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-
Subtotal	3	-	-	-			
Figural Information							
17	6	4	4	4	18	0.61	0.28
18	2	-	5	6	13	0.06	-0.03
19	2	-	5	6	13	0.11	0.08
20	6	4	5	3	18	0.57	0.19
21	6	4	6	2	18	0.65	0.31
22	6	4	4	2	16	0.60	0.22
23	-	-	1	2	3	0.48	0.37
24	1	1	2	-	4	0.08	0.15
25	-	-	-	-	-	-	-
26	-	-	2	2	4	0.65	0.30
27	-	-	2	2	4	0.45	0.24
Subtotal	29	17	36	29			
Reading Load							
28	-	3	1	1	5	0.66	0.26
29	5	2	6	4	17	0.55	0.21
30	4	-	2	4	10	0.31	0.41
31	1	2	-	4	7	0.53	0.49
32	11	11	8	7	37	0.06	0.18
33	-	-	-	-	-	-	-
34	-	1	-	-	1	0.96	0.91
Subtotal	19	19	17	24			
Process Skills							
35	-	-	-	4	4	0.29	0.22
36	-	-	-	4	4	0.29	0.22
37	-	-	-	4	4	0.04	-0.06
38	-	-	-	-	-	-	
Subtotal	-	-	-	12	-		

Table 4. Some Characteristics of the Grade 4 Blocks Used in the Study

Block Type	Item Types				Approximate N[1]	Mean p-values	Mean % Not Reaching Items[2]	% Not Reaching Last Item
	Total	Multiple Choice	Constructed Resp. – D[3]	Constructed Resp. – P[4]				
Conceptual/PS 1	11	6	2	3	1200	51	2.5	13.2
Conceptual/PS 2	11	6	-	5	1200	47	4	16.9
Theme (Life Cycle)	8	1	1	6	1600	47	6.8	34.2
Hands-on Task (Floating Pencil)	7	2	-	5	1800	44	15	47.2
Totals	**37**	**15**	**3**	**19**				

[1] The p-values presented in the table were obtained with nationally representative samples ranging from about 1200 to 1600, depending on the block type, however the sample size for the application of the rule-space methodology was about 160 for each of the conceptual/problem-solving blocks and about 320 each for the theme block and the task.

[2] Subjects not reaching a given item are excluded from computation of item statistics.

[3] D = Dichotomous – these are constructed-response items that are scored either right or wrong.

[4] P = Polytomous – these are constructed-response items with one or more partial credit levels.

Figure 2. Rule Space Analysis of Four Blocks in the 1996 NAEP Science Assessment Grade 4 -- Progress Paths 1-3

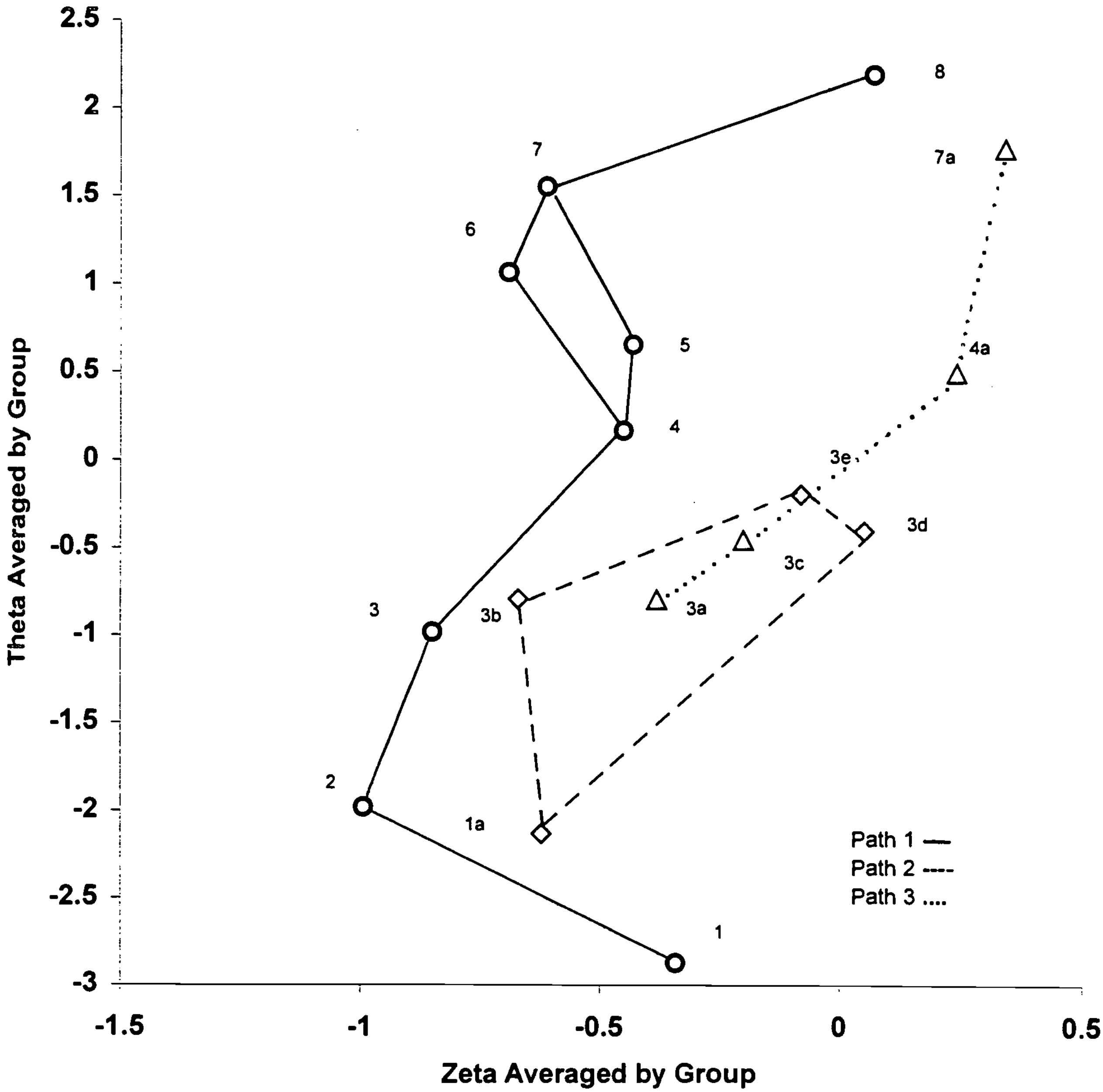


Table 5. Knowledge States for Progress Path 1 for Two Books of the 1996 NAEP Science Assessment - Grade 4

State	Mean Score	N	Attributes Mastered						Description of Mastery Levels
			Knowl	Reason	Hypoth	Figural	Read Load	Process	
1	-2.87	5	1-3, 5-6			17, 20-22, 26	28-29		**K1** (Knowledge 1) = 1-3, 5-6 **F1** (Figural 1) = 17, 20-22, 26 **L1** (Reading Load 1) = 28-29
2	-1.98	6	1-6			17, 20-22, 26	28-29		**K2** (Knowledge 2) = K1 + 4 (principles) F1 L1
3	-1.01	14	1-6	8		17, 20-23, 26-27	28-29		K2 **R1** (Reasoning 1) = 8 (cause-effect) **F2** (Figural 2) = F1+ 23, 27 (stu-gen info) L1
4	+0.17	32	1-6	7		17, 20-23, 26	28-31		K2 **R2** (Reasoning 2) = 7 (deductive) F2 (minus 27) **L2** (Reading Load 2) = L1 + 30-31
5	+0.66	8	1-6	7		17, 20-23, 26-27	28-31	35-36	K2 R2 F2 L2 **P1** (Process1) = 35-36 (manip & record)
6	+1.07	9	1-6	7, 11		17, 20-23, 26	28-31		K2 **R3** (Reasoning 3) = R2 + 11 (models) F2 (minus 27) L2 No P1

7	+1.56	16	1-6	7, 11		17, 20-23, 26-27	28-31	35-36	K2 R3 F2 L2 P1
8	+2.20	4	1-6	7, 10-12	14	17, 20-23, 26-27	28-31	35-36	K2 **R4** (Reasoning 4) = R3 + 10, 12 (explan) **H1** (Hypothesis 1) = 14 (variable identif) F2 L2 P1

Table 6. Knowledge States for Progress Path 2 for Two Books of the 1996 NAEP Science Assessment - Grade 4
*** Indicates Path I Stage**

State	Mean Score	N	Attributes Mastered						Description of Mastery Levels
			Knowl	Reason	Hypoth	Figural	Read Load	Process	
1*	-2.87	5	1-3, 5-6			17, 20-22, 26	28-29		**K1** (Knowledge 1) = 1-3, 5-6 **F1** (Figural 1) = 17, 20-22, 26 **L1** (Reading Load 1) = 28-29
1a	-2.13	11	1, 4-6	10, 12			34		**K1A** = 1, 4-6 **R1A** 10, 12 (explanation) **L1A** = 34
3b	-0.79	27	1, 4-6	7, 10, 12			34		K1A **R2A** = R1A + 7 (R3) L1A
3d	-0.40	17	1, 4-6	10-12			31-34		K1A R1A **L2A** = L1A + 31
3e	-0.19	35	1, 4-6	7, 10, 12			31-34		K1A R2A L2A
6*	+1.07	9	1-6	7, 11		17, 20-23, 26	28-31		**K2** (Knowledge 2) = K1 + 4 (principles) **R3** (Reasoning 3) = R2 + 11 (models) **F2** (minus 27) **L2** (Reading Load) = L1 + 30-31
7*	+1.56	16	1-6	7, 11		17, 20-23, 26-27	28-31	35-36	K2 R3 F2 L2 **P1** (Process1) = 35-36 (manip & record)

8*	+2.20	4	1-6	7, 10-12	14	17, 20-23, 26-27	28-31	35-36	K2 **R4** (Reasoning 4) = R3 + 10, 12 (explan) **H1** (Hypothesis 1) = 14 (variable identif) F2 L2 P1

Table 7. Knowledge States for Progress Path 3 for Two Books of the 1996 NAEP Science Assessment- Grade 4
*** Indicates Path I Stage**

State	Mean Score	N	Attributes Mastered						Description Of Mastery Levels
			Knowl	Reason	Hypoth	Figural	Read Load	Process	
1*	-2.87	5	1-3, 5-6			17, 20-22, 26	28-29		**K1** (Knowledge 1) = 1-3, 5-6 **F1** (Figural 1) = 17, 20-22, 26 **L1** (Reading Load 1) = 28-29
2*	-1.98	6	1-6			17, 20-22, 26	28-29		**K2** (Knowledge 2) = K1 + 4 (principles) F1 L1
3a	-0.80	9	1-6			17, 20-23, 26-27	28-29	35-36	K2 **F2** (Figural 2) = F1 + 3, 27 (stu-gen info) L1 **P1** (Process1) = 35-36 (manip & record)
3c	-0.46	17	1-6	8		17, 20-23, 26-27	28-29	35-36	K2 **R1** (Reasoning 1) = 8 (cause-effect) F2 L1 P1
4a	+0.50	7	1-6	7-8		17, 20-23, 26-27	28-31		K2 R1 **R2** (Reasoning 2) = 7 (deductive) F2 **L2** (Reading Load 2) = L1 + 30-31 P1
7a	+1.78	3	1-6	7-8 11		17, 20-23, 26-27	28-31	35-36	K2 R1 **R3** (Reasoning 3) = R2 + 11 (models) F2 L2 P1

8*	+2.20	4	1-6	7, 10-12	14	17, 20-23, 26-27	28-31	35-36	K2 **R4** (Reasoning 4) = R3 + 12 (explanation) **H1** (Hypothesis 1) = 14 (variable identif) F2 L2 P1

Figure 2. Graphical Representation of Progress Path 1 for Two Books of the 1996 NAEP Science Assessment -- Grade 4

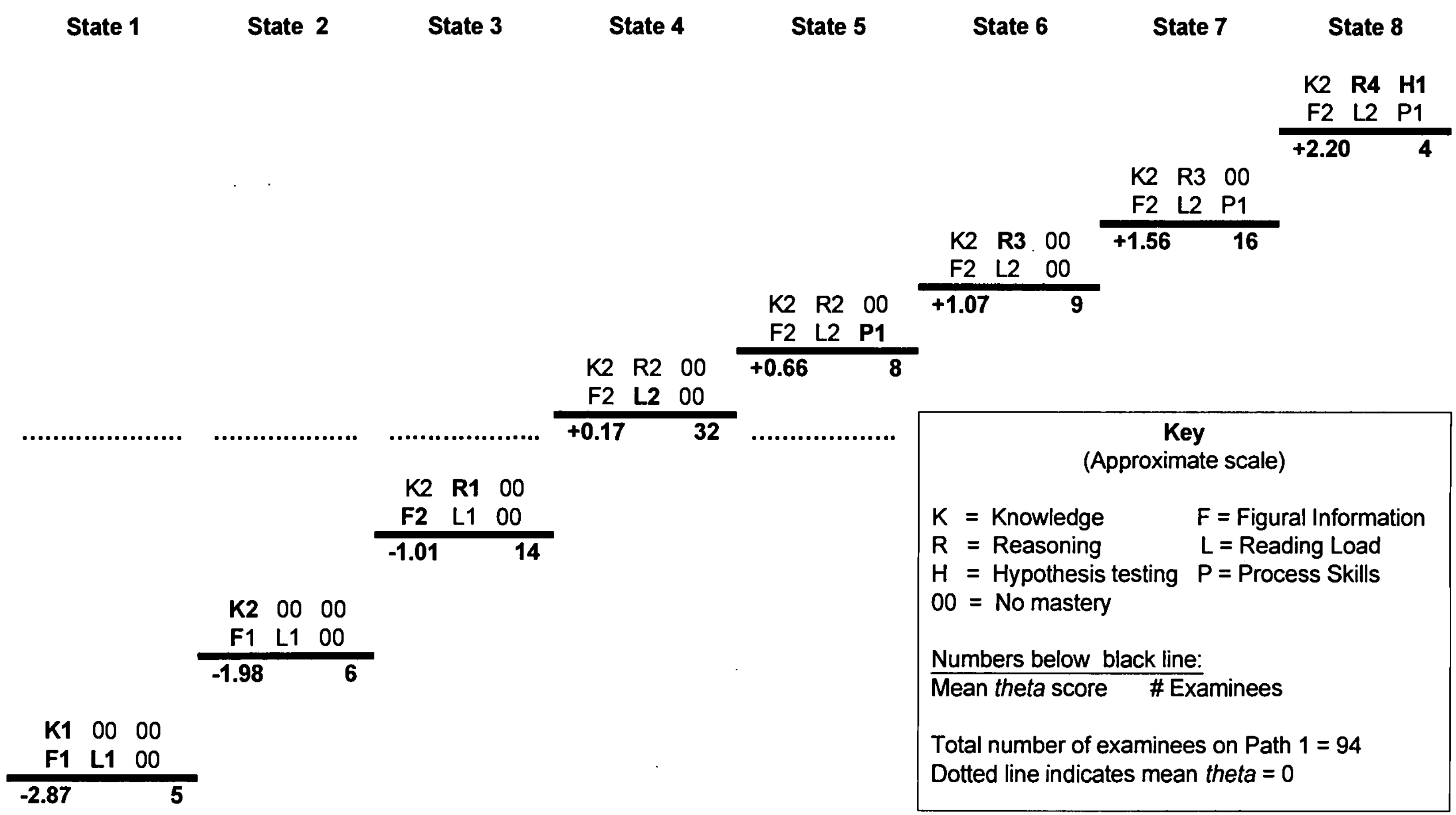


Figure 3. Graphical Representation of Progress Path 2 for Two Books of the 1996 NAEP Science Assessment - Grade 4

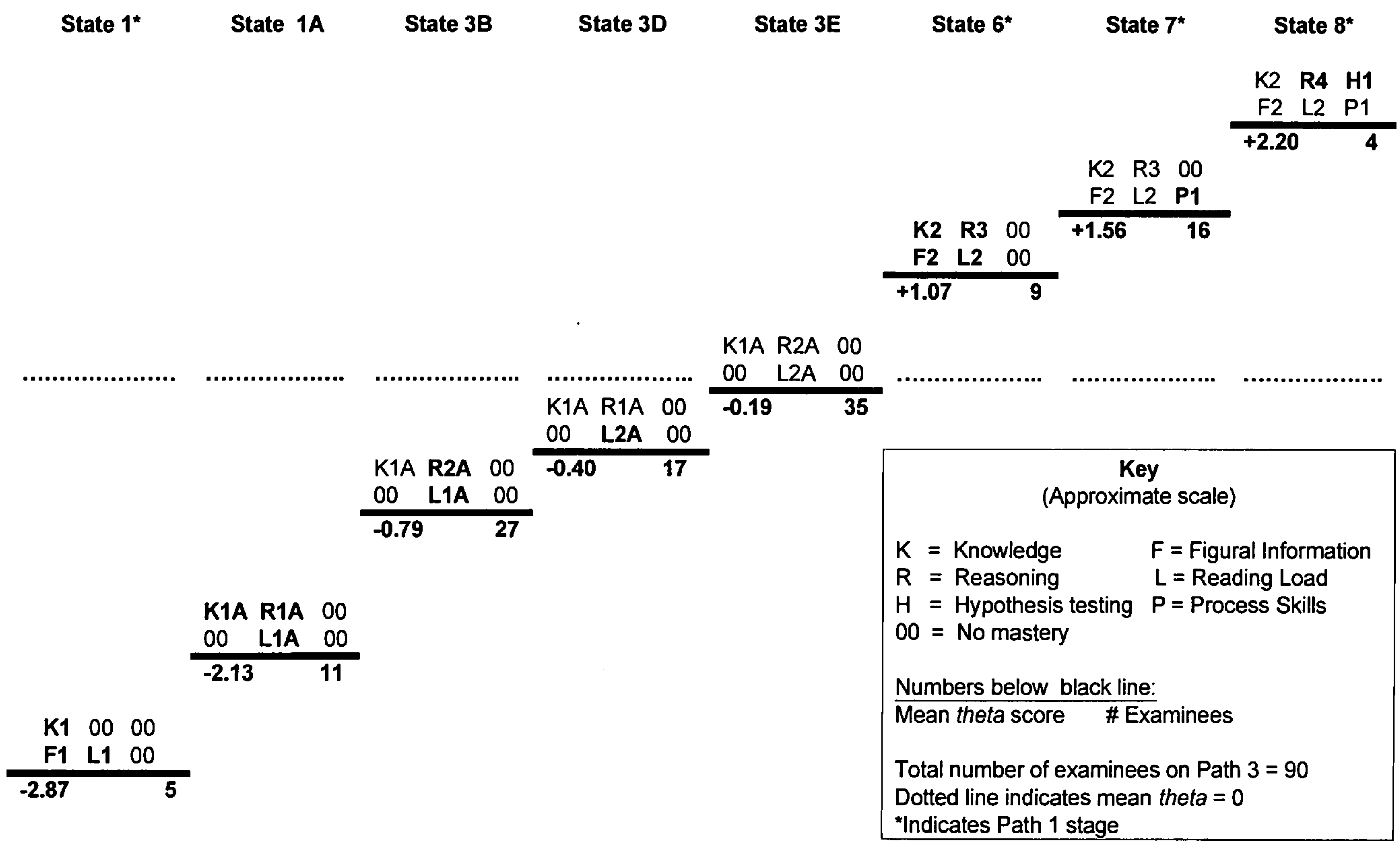
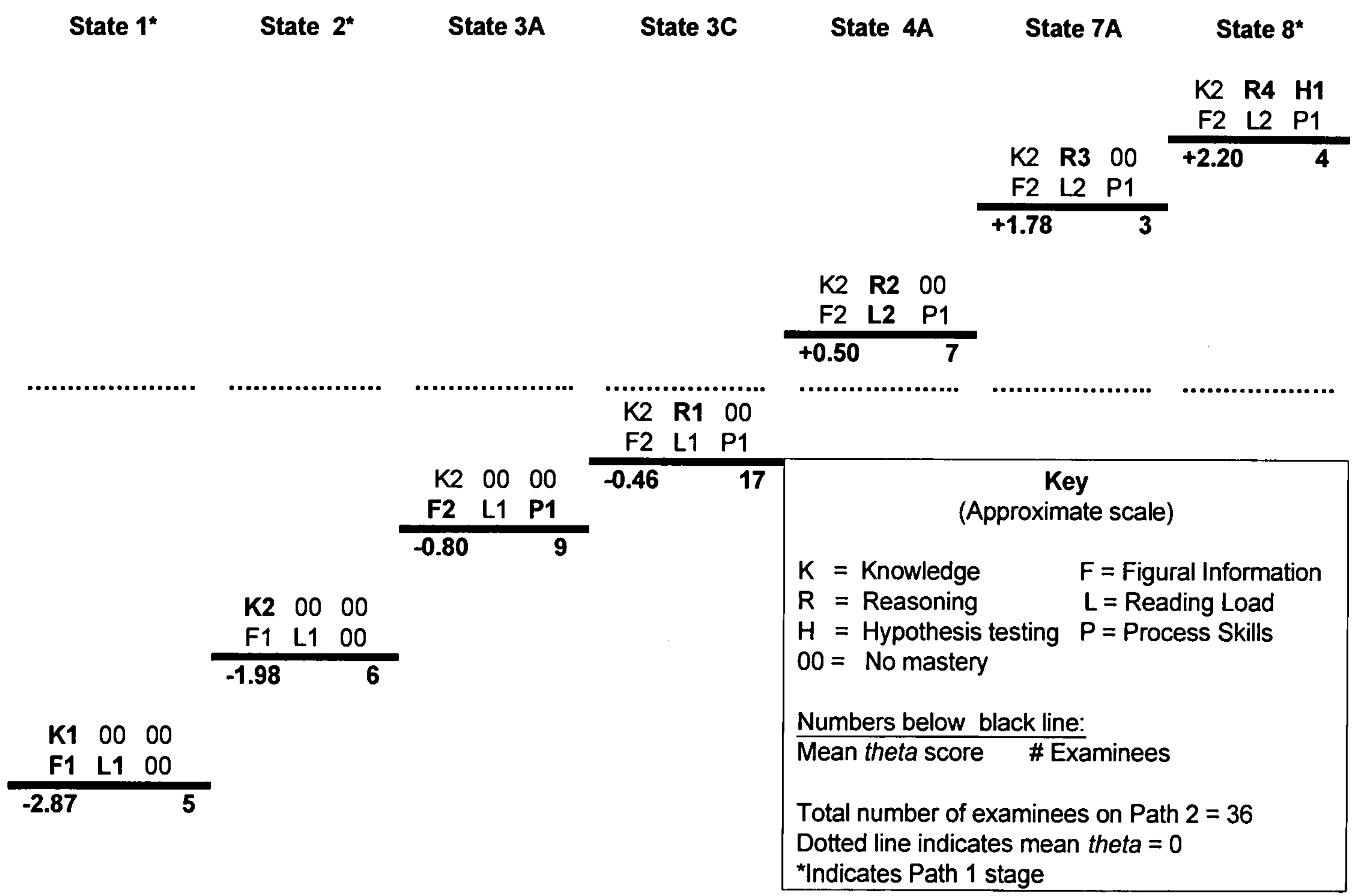


Figure 4. Graphical Representation of Progress Path 3 for Two Books of the 1996 NAEP Science Assessment - Grade 4



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM028945

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Application of the Rule-Space Methodology to the 1996 NAEP Science Assessment: Grade 4 Preliminary Results

Author(s): Mario Yepes-Baraya, Ph.D.; And Others

Corporate Source:	Publication Date:
Educational Testing Service	April 1998

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☒	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]	Printed Name/Position/Title: Mario Yepes-Baraya, Associate Research Scientist	
Organization/Address: Educational Testing Service (8-R), Princeton, NJ 08541	Telephone: (609)734-5357	FAX: (609)734-1755
	E-Mail Address: myepes@ets.org	Date: April 23, 1998

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com