DOCUMENT RESUME

ED 388 676 TM 023 675

AUTHOR Yepes-Baraya, Mario
TITLE Task Analysis of Science Performance Tasks and Items: Identifying Relevant Attributes.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE Apr 95
NOTE 24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Processes; Educational Assessment; Elementary Secondary Education; Field Tests; National Surveys; Protocol Analysis; Questionnaires; Research Methodology; Science Instruction; Sciences; *Science Tests; *Task Analysis; *Test Construction; Test Validity
IDENTIFIERS National Assessment of Educational Progress; *Performance Based Evaluation

ABSTRACT

This paper describes the task analysis of performance-based science tasks that were designed for the 1994 National Assessment of Educational Progress (NAEP) science assessment, now postponed until 1996, and field tested in 1993. A brief description of the science performance tasks is followed by a description of the task analyses performed and a discussion of the various uses of task analysis. In 1993, 17 performance-based tasks were field tested, with 5 for grade 4, 2 for grades 4 through 8, 3 for grade 8 only, 2 for grades 8 through 12, and 5 for grade 12 alone. In level one task analysis, the unit of analysis was the task as a whole. Analysis methods included observations, the think aloud method, questionnaires, and expert opinions. In level two analysis, the cognitive operations and skills of steps in the task were analyzed. As results in these field tests demonstrated, level one analysis results in a mapping of the task's structure from the perspectives of test takers. Level two analysis shows the nature of the skills for completing the task. Together, the levels can ascertain the extent to which tasks meet standards of validity. (Contains 3 tables, 2 figures, and 13 references.) (SLD)

Task Analysis of Science Performance Tasks and Items: Identifying Relevant Attributes

Mario Yepes-Baraya

Division of Cognitive and Instructional Science
Educational Testing Service
Princeton, NJ 08541

Paper presented as part of the Symposium *Large-Scale Science Performance Assessment Results: Informing Test and Score Development* at the annual meeting of the National Council for Measurement in Education, San Francisco, April 1995.

Task Analysis of Science Performance Tasks and Items: Identifying Relevant Attributes

Mario Yepes-Baraya
Educational Testing Service, Princeton, NJ 08541

Introduction

This paper describes the task analysis of performance-based science tasks that were designed for the 1994 National Assessment of Educational Progress (NAEP) science assessment, now postponed until 1996, and field tested in January and February 1993. A brief description of the science performance tasks is presented first, followed by a description of the task analyses performed, and a discussion of the various uses of task analysis.

NAEP's Performance-Based Tasks

As shown in Table 1, seventeen performance-based tasks were field tested in 1993: five were grade 4 only tasks, two were grade 4/8 overlap tasks, three were grade 8 only tasks, two were grade 8/12 overlap tasks, and five were grade 12 only tasks. Whereas the grade 8 and grade 12 versions of the two grade 8/12 overlap tasks were identical, this was not true for the two grade 4/8 overlap tasks. For both of these tasks, the grade 8 version is slightly different and longer than the grade 4 version. As a result of these differences, there were nineteen uniquely different tasks.

The introduction to the NAEP science assessment tasks conform to one general format. Each task's first page has the name of the task, the instructions presented below, and a diagram of the equipment and materials contained in the plastic bag given to each student being assessed. The instructions read as follows:

> For this task, you have been given a kit that contains materials that you will use to perform an investigation during the next 20 minutes [30 minutes for

grade 8 and grade 12 tasks]. Please open your kit and use the following diagram to check that all the materials in the diagram are included in your kit. If any materials are missing, raise your hand and the administrator will provide you with the materials that you need.

On the second page, after describing the purpose of the investigation and before the first step in the task, students are told to "... follow the directions step-by-step and write the answers to the questions in the space provided in your booklet". Beyond the introductory information there is a series of steps for students to follow, ranging from 4 to 12, depending on the task. Most steps within a task correspond to discrete items requiring students to answer a question or provide some information. Some steps within a task, however, do not correspond to a discrete item; their function is simply to provide instructions or information. Only those steps corresponding to discrete items are scored. As described by O'Sullivan (1995), there are three basic types of tasks with respect to the way the directions and the questions are scaffolded, ranging from more structured to less structured. The first type of tasks requires that students write down responses as they proceed through each step. In the second. type students are first instructed to perform certain activities and then asked to answer some questions. The third type presents a problem for students to solve given the equipment and materials contained in the kit.

The format of the items that are part of a task falls into two broad categories: multiple choice and constructed response. Constructed response items can in turn be classified as short, long, or extended depending on the length/complexity of the expected response.

The above description of NAEP's performance-based tasks is not so different from the more general description of a task offered by Romiszowski (1986) in the context of designing instructional systems:

> A task is a coherent set of activities (steps, operations, or behavior elements) which leads to a measurable end or result. The steps of a task are therefore interrelated.

Task Analysis of Science Performance Tasks and Items

As used here, task analysis refers to two levels of analysis: Level One and Level Two. For Level One the unit of analysis is the performance task as a whole. For Level Two the unit of analysis is each of the items that are part of a given performance task. Each of these levels is discussed below.

Level One Analysis. According to Romiszowski (1986), the task analysis of a performance task results in a set of steps to be performed and one or several operational sequences, depending on how the task is structured. As suggested above, in the case of the NAEP's performance tasks, in most cases the steps of the task are specified by the step-by-step instructions provided. The operational sequence, however, is more complex than simply following steps *1* through *n*. What is of interest to us in determining the operational sequence(s) of a performance task is to map out how respondents process the information necessary to work through the task. When this is done, the interrelationships among the tasks' steps should be apparent. Level One analysis methods include observations, the think aloud method, interviews, questionnaires, and expert opinion (Anderson, Ball, & Murphy, 1975; van Someren, Barnard, & Sandberg, 1994). For the study reported here, a combination of observations, interviews, and expert opinion was used.

Figure 1 illustrates the results of a Level One analysis for CANDY, one of the fourth-grade performance tasks that was field tested but not selected for the 1996 assessment. For this task (see transparencies), students are given a small packet of white powder and three pieces of candy, each a different color. Students are first given detailed instructions to perform a test on the white powder with sugar-testing paper, and to record what they observe. If the paper changes color, they are told, then it must have had sugar in it. Students are then asked to use the materials in their kit to test the three pieces of candy to find out which ones were made with sugar and which one were not. After they have tested the three pieces of candy, they are asked three open-ended questions. The first question is about the process they followed to determine whether the pieces of candy had sugar or not. The second question asks for the results obtained and the evidence for these results. The third question presents the results obtained on a similar test with a piece fruit in which the paper changed color, and then asks what can be concluded about the piece of fruit.

On Figure 1 the larger size letters in capitals identify stimuli or information provided. *K* is the kit of materials and equipment as well as the first page of the test which provides labels for the kit elements. *A1* is the detailed information given on pages two and three (steps a-e) about testing for sugar. *Q* is a question on page 3 with a yes/no answer. The question is not a test item and therefore is not scored. The question is metacognitive, i.e., its purpose is to make students aware of how the sugar-testing paper works. *A2* is the instructions given at the top of page 4 to test the three pieces of candy. The letters in small case, *a1* and *a2*, represent the student's carrying out the operations indicated by *A1* and *A2*. *A2* is enclosed in a diamond to indicate that unlike *A1*--in which students are told in detail what to do--students must generate the information to perform *a2*.

Roman numerals identify the task's steps. Circles around numerals indicate test items, i.e., steps requiring students to answer a question or provide information. Squares around circled numerals identify individually scored items.

Arrow lines identify the processing of information as students perform the task. Solid arrow lines indicate dependency relationships between operations performed, between an operation and the response to an item, or between items. For example, $X \longrightarrow Y$ means that *X* must be done successfully in order for *Y* to be done successfully. In other words, *Y* is dependent on *X*. Dotted arrow lines identify contingency relationships between elements in the task.

X ----> *Y* means that the information in *X* might be helpful (but is not necessary) for doing *Y* successfully.

Level One analysis of the CANDY performance task reveals the following:

- As shown by the solid arrow lines, successful performance on operation *a1* directly depends on the test takers' ability to understand and carry out instructions *A1*. It also depends on their ability to handle and/or familiarity with manipulating similar equipment and materials.

- As shown by the solid arrow lines, successful performance on operation *a2* directly depends on the test takers' ability to understand and carry out instructions *A2*. However, since *A2* depends on *A1*, *a2* is also dependent on *A1*.

- The dotted arrow lines between *a1* and *a2* suggest that performing *a1* successfully might be helpful in performing *a2*. However, it is not necessary. The same relationship exists between *a1* and *Q*.

- Item *1*, as stated above, asks for the process followed to determine whether the pieces of candy had sugar or not. The item is an extended constructed-response item. A fairly detailed description is necessary to obtain full credit. The dotted arrow lines between *A1* and *1* and between *a2* and *1* suggest that reading/copying the instructions given in A1 might be as helpful in answering the item as reconstructing the steps actually followed in performing the tests with the three pieces of candy.

- Item *2* has two parts to it. The first part (*2a*) asks for an interpretation of the results obtained in testing the three pieces of candy (i.e., whether the candy has sugar or not). As suggested by the solid arrow line, successful performance in *2a* is solely dependent on correctly interpreting the information given in *Q*. Doing *a2* correctly might be helpful, but it is not necessary. The second part (*2b*) asks for the evidence for the conclusion arrived at in *2a*. As suggested by the solid arrow line, successful performance in *2b* depends on doing *a2* correctly.

- Item *3* presents the results obtained on a similar test with a piece fruit in which the paper changed color, then asks what can be concluded about the piece of fruit. The dotted arrow lines suggest possible relationships with other sources of information.

- Overall, it appears that *A1* (the instructions to perform the sugar test) and *Q* (how to interpret the results) are the critical sources of information for CANDY. *A1* is one of the paths for item *1*. *Q* is required for item *2a* and is one of the paths for item *3*. Item *2b* is the only item that requires test takers to actually do the tests (*a2*). However, *a2* is also a possible path for items *1*, *2a*, and *3*.

Each of the NAEP tasks has a different structure, as revealed by a Level One analysis. Figure 2 illustrates the task analysis of another grade four performance task somewhat more complex than CANDY.

Level Two Analysis. As one performs a Level One analysis on a performance task, such as the one described for CANDY above, one becomes aware not only of the task's operational sequence(s), (i.e., how the information in one step/item is related to the information in the other steps and/or items) but also of the type of cognitive operations/skills required to perform each step and answer each item. Some items, for example, ask questions designed to assess *comprehension* of a complex set of data. Other items require that test takers *record* their observations as they perform a task. Yet, other items may require test takers to access information not given on the test but learned in everyday life or in the course of formal science instruction, e.g., to *recall* facts, *explain* concepts, or *apply* principles.

By focusing on the individual steps and items that make up a task, the objective of a Level Two task analysis is to identify the cognitive operations/skills that test takers use in performing each step and answering each item. A review of the literature suggests several sources of cognitive operations/skills categories for science assessments (Yepes-Baraya & Allen, in preparation). For the Level Two analysis, analysts use coding questions that match the cognitive operations/skills that items may require. Table 2 presents a list of thirty-six coding questions, arranged in seven main categories, used in doing a Level Two analysis for six blocks of items in the 1993 NAEP science field test. In addition to items belonging to two performance-based tasks, the task-analyzed items also included the items that were part of two theme blocks (in which all items pertain to a common theme, e.g., the solar system), and the items in two regular blocks (in which items are usually unrelated to one another). If an item requires test takers to use a certain cognitive operation/skill, then the item is said to possess a corresponding feature or attribute. For example, if an items contains science terminology/ vocabulary that must be understood in order to answer the item correctly (e.g., the word . "omnivore"), the item possesses that attribute (Attribute 17 in Table 3). Table 3 illustrates the coding of each of the four items in CANDY with each of the thirty-six attributes identified in the NAEP science study.

Various Uses of Task Analysis

Test development, like teaching, is both an art and a science. Despite the abundant literature on test design (and instructional design, as well), test developers usually approach the design of a new test intuitively, especially if they are experienced professionals. The design of performance-based science tasks, like the nineteen unique tasks developed for the NAEP science assessment, is a recent phenomenon prompted by the call for more authentic modes of assessment that tap into students higher-order thinking skills. After all, scientists in the real world spend extended periods of time conducting complex experiments and activities. Unlike more traditional performance tasks used in earlier large-scale assessments (for example those used in the Second International Science Study; Kanis, Doran & Jacobson, 1990), the NAEP tasks consist of more than just one or two steps, are more complex and cognitively demanding, and take longer to complete. All of these tasks were first pilot tested with groups of students in the classroom, and then field tested with nationally representative samples. Nevertheless, the task analysis of these tasks, prior to or concurrent with the pilot and field testing can be helpful in improving test design.

What can be learned from Level One and Level Two analyses? As shown with the example of CANDY, a Level One analysis results in a mapping of a task's structure from the test takers' perspectives. A task's structure may reveal unintended consequences of the task. For example, a given task may rely too heavily on a particular key step upon which several other steps depend, and the step in question may in turn depend excessively on reading comprehension instead of science knowledge and/or reasoning skills.

Because these are timed tasks, an estimate of the time necessary to complete each step (t_1, t_2, ..., t_n) can be made in order to determine the degree to which the task is speeded. In the case of CANDY, if the total time is 20 minutes, and if $t\ (A1+a1) = 5$ minutes, $t\ (A2+a2) = 12$ minutes, and $t\ (1) = 2\text{-}3$ minutes, then we could predict that some students might not have enough time to do *2a*, *2b*, and *3*.

In combination with a Level Two analysis, a Level One analysis can be used to modify a given task. In the case of CANDY, for example, the Level Two analysis shows the nature of the skills that are part of the critical path for completing the task successfully, as well as the skills that are not part of the critical path. If a given performance task were too speeded and needed to be shortened, results of the task analyses might be useful in helping to decide how to best accomplish this end.

In general, Level One and Level Two analyses can be used to ascertain the extent to which tasks meet the uniform standards of validity (Messick, December 1994) which apply to all educational and psychological assessments, including performance assessments. Given that NAEP is essentially a construct-based assessment (as opposed to following a task-centered approach to task development), task analysis and the NAEP framework (NAEP Science Consensus Project, 1993) can serve as the basis to identify sources of construct-irrelevant variance and/or construct underrepresentation. In the case of the science performance tasks, it is possible that the simultaneous demands to measure higher-order thinking while providing specific directions to facilitate task completion within the allotted time resulted in both a narrowing of the construct to be assessed (Enright, November 1993) and the introduction of ancillary processes which are a potential source of construct-irrelevant variance.

In addition to improving test design and enhancing the validity of the assessment, task analysis of performance-based tasks can also be used to modify or improve the scoring of these tasks. This analysis is particularly applicable to fairly structured tasks than can be improved by making them more procedurally open-ended (Baxter, Glaser & Raghavan, 1994). It is useful when, instead of scoring the steps of the task individually, a holistic scoring rubric would be more appropriate.

Level Two analysis has also been used as the basis for statistical analyses of the sort that Tatsuoka (1993, 1990), Yepes & Allen (in preparation) and Park & Allen (1994) have reported, where item attributes or clusters of these attributes are used as predictors of item difficulty.

Notes

This project is a part of an ETS research program to understand what the next NAEP science assessment will measure and to develop potential ways to summarize the results for that assessment. It was funded with ETS research allocation funds. The author acknowledges with gratitude the continued support of Nancy L. Allen of the Division of Psychometrics and Statistics at ETS. Opinions expressed are those of the author and not necessarily ETS or NAEP.

All correspondence should be sent to

Mario Yepes-Baraya, ETS (08-R), Princeton, NJ 08541,

or to myepes@ets.org

References

Anderson, S.B., Ball, S., & Murphy, R.T. (1975). *Encyclopedia of educational evaluation.* San Francisco: Jossey Bass.

Baxter, G., Glaser, R. & Raghavan, K. (1994). *Analysis of cognitive demand in selected alternative science assessments.* (CSE Technical Report 382). Los Angeles: UCLA's Center for the Study of Evaluation.

Enright, M. (November 1993). *Approaches to performance assessment task development.* (Working Seminar Report WS1). Princeton, NJ: Center for Performance Assessment, Educational Testing Service.

Kanis, I.B., Doran, R.L., & Jacobson, W.J. (1990). *Assessing science laboratory skills at the elementary and middle/junior high levels -- second international science study (SISS).* New York: Teachers College, Columbia University.

Messick, S. (December 1994). *Alternative modes of assessment, uniform standards of validity.* (Research Report RR-94-60). Princeton, NJ: Educational Testing Service.

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.

O'Sullivan, C. (1995). *The 1993 NAEP science field test: Hands-on tasks and test specifications.* Paper to be presented at the annual meeting of the National Council for Measurement in Education, San Francisco.

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 the American Educational Research Association, New Orleans, LA.

Romiszowski, A.J. (1984). *Designing instructional systems: Decision-making in course planning and instructional design.* New York: Nichols Publishing.

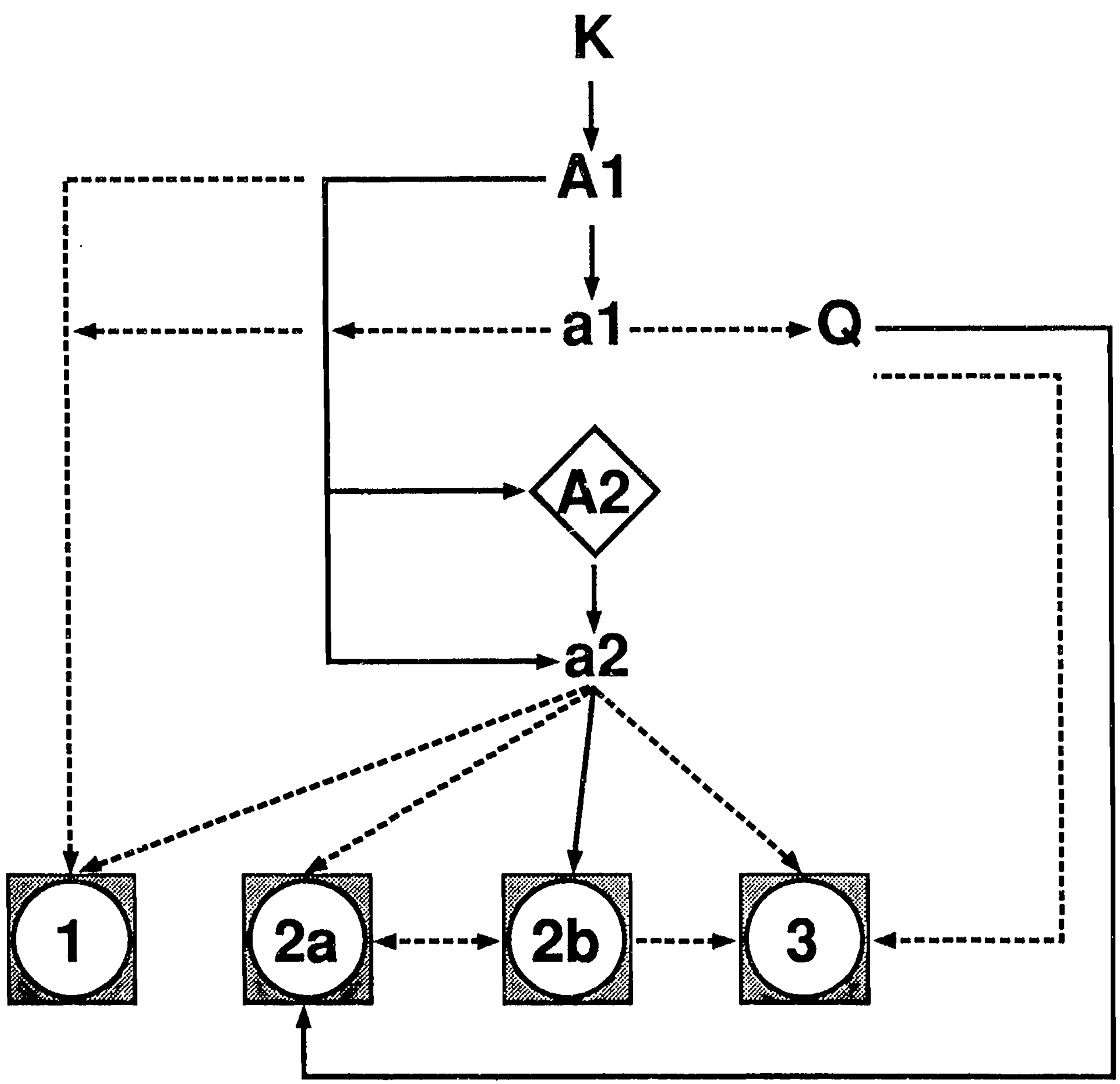
Tatsuoka, K.K., Birenbaum, M., Lewis, C., Sheehan, K. (1993). *Proficiency scaling based on conditional probability functions for attributes.* (Research Report RR-93-50-ONR). Princeton, NJ: Educational Testing Service.

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.

van Someren, M.W., Barnard, Y.F., Sandberg, J.A.C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes.* San Diego: Academic Press.

Yepes-Baraya, M., & Allen, N.L. (In preparation). *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.

FIGURE 1. Level One Task Analysis of CANDY



Legend

K, A1, Q, A2	Sources of information for test taker
a1, a2	Operations performed by test taker
1, 2a, 2b, 3	Task steps
◇	Instructions requiring test taker to generate information
○	Test item
■ (shaded square)	Individually scored item
———→	Dependency relationships
- - - -→	Contingency relationships

FIGURE 2. Level One Task Analysis of a Grade Four Performance Task

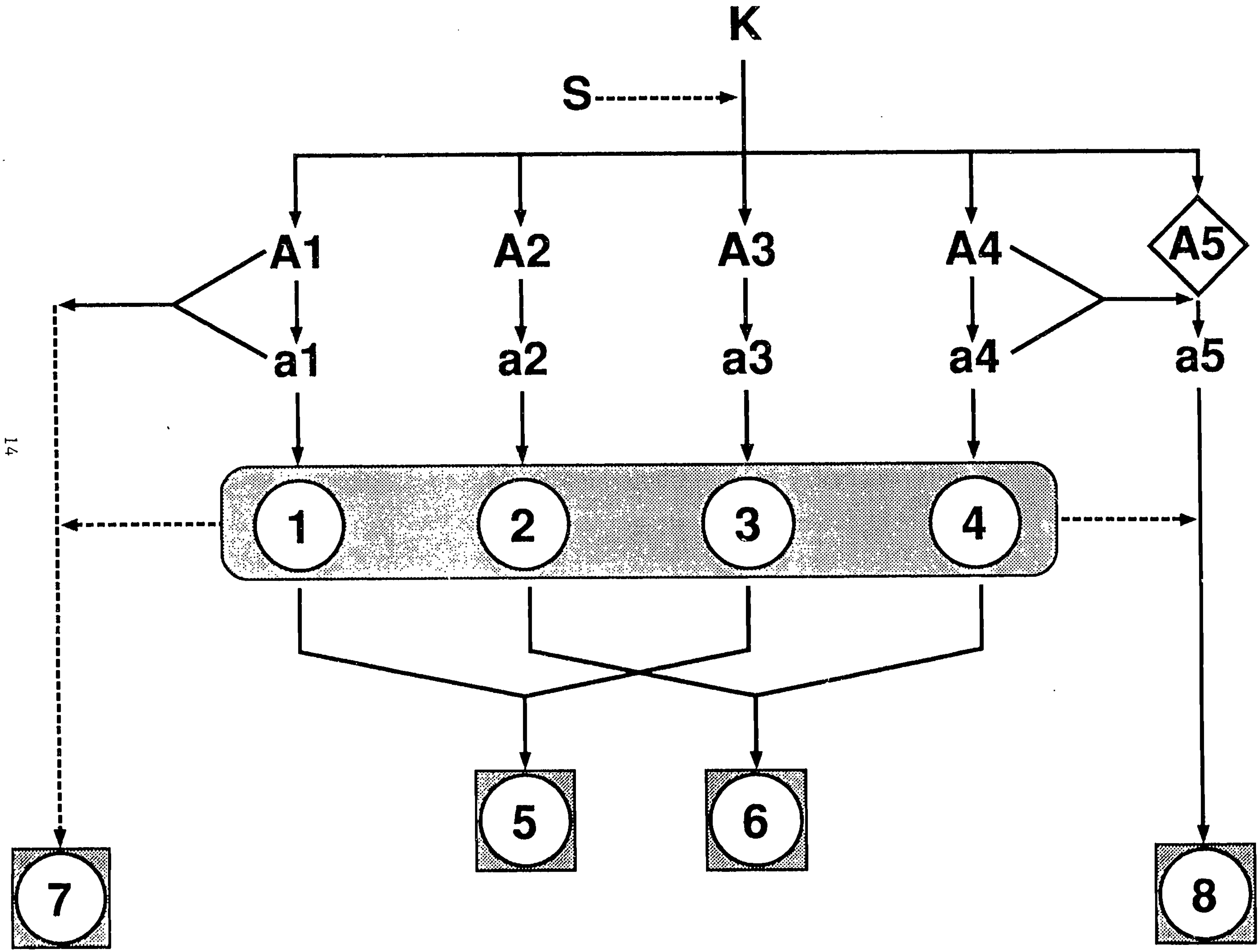


TABLE 1. 1993 NAEP Science Field Test

PERFORMANCE TASKS BY GRADE LEVEL

Key: H = Hands-On Performance Tasks

Booklet #	Grade Level	Performance Tasks	
1	4	H1	
2	4	H2	
3	4	H3	
4	4	H4	
5	4	H5	
6/7	4/8	H6A/B	4-8 OVERLAP BLOCKS
8/9	4/8	H7A/B	
10	8	H8	
11	8	H9	
12	8	H10	
13/14	8/12	H11	8-12 OVERLAP BLOCKS
15/16	8/12	H12	
17	12	H13	
18	12	H14	
19	12	H15	
20	12	H16	
21	12	H17	

TABLE 2. Coding Questions that Match the Skills that Items May Require[1]

A. Knowledge

In each case knowledge is brought to the task by the test taker, as opposed to being provided in the item. The general format of the coding questions in the knowledge category is:

Can knowledge of _______ be used to answer the item?

1. Facts
2. Experimental procedures
3. Concepts
4. Principles
5. Relationships

B. Information in a Text, Table, Graph, or Figure (tTGF)

In each case reference is made to information in

- a TGF = table, graph, or figure, OR
- a tTGF = text, table, graph, or figure

The stimulus material is counted separate from the item.

The TGF/tTGF can be

- given (g), OR
- student-generated (s).

6. Does the item contain a TGF that is already completed or needs to be completed?
7. Does the item refer, directly or indirectly, to the information contained in a completed TGF?
8. Does the item refer, directly or indirectly, to the information in a tTGF (s) separate from the stem?

---

[1]Adapted from *The Process of Identifying Item Attributes Related to Item Performance for the 1993 National Assessment of Educational Progress Science field Test* (Yepes-Baraya & Allen, in preparation).

TABLE 2. Coding Questions that Match the Skills that Items May Require (Continued)

9. Does the item contain (or refer to the information contained in) a completed TGF (g or s) that is complex?

10. When present, is it possible to use information contained in the completed TGF (g or s) to answer the item?

11. Is it necessary to use information contained in the completed TGF (g or s) to answer the item no matter which strategy is used?

12. Is some of the information needed to answer the item contained in a TGF (s)?

13. Is all the information needed to answer the item available in the tTGF in the block with the item? [All information is (g)]

14. Is all the information necessary to answer the item available in the tTGF in the block with the item? [Some of the information is (s). Information does not include procedural knowledge, which may be needed in addition to the information provided in the block of items]

15. Does the response require a TGF to be drawn or completed?

16. Does the response require a GF to be drawn or completed?

C. Information in the Item Itself

17. Does the item contain science terminology/vocabulary that must be understood in order to answer the item correctly?

18. Must the response meet all the conditions specified in the stem?

19. Does the item contain hypotheticals, suppositions, exceptions, or negations (that make the task complex)?

20. Can the item be solved by choosing the odd option out?

21. Does the item require information that could have been gained through practical experience (not formally instructed)?

22. Does the item require only information found in the item itself (including procedural knowledge)? [Not (s)]

TABLE 2. **Coding Questions that Match the Skills that Items May Require** (Continued)

D. Reasoning

23. Can reasoning from a general concept, principle or law to a specific conclusion be used to answer the item?

24. Is tracing a cause-effect from one component to another within a system necessary to answer the item?

25. Can formal inductive reasoning be used to answer the item?

26. Can the application of a concept, principle, or general idea be used to answer the item?

27. Can thinking with models/analogies (or thinking visually) be used to answer the item? [The model must not be present, but be brought in by student]

E. Hypothesis Generation/Testing

28. Is the generation of a hypothesis or prediction necessary to answer the item?

29. Does the item require the identification of variables and/or a control group involved in a design of a test for a hypothesis?

30. Does the item require the generation of specifically operationalized procedures to be used in testing a hypothesis?

31. Does the item require the use of multiple control groups in a design of a test for a hypothesis?

F. Explanation

32. Does the item require that a response be given and that the response be justified?

G. Item Format

33. Is the item a 4-category, extended constructed-response item?

34. Is the item a short constructed-response item?

35. Does the stem of the item have one or more intratext referentials (e.g., it, this, these)?

36. Does the stem of the item have one or more clauses with fronted structures?

TABLE 3. Coding of Items in CANDY

(36 coding attributes were used)

Item # → Attributes ↓	1	2a	2b	3
1. Can knowledge of facts be used to answer the item?	0	0	0	0
2. Can knowledge of experimental procedures be used to answer the item?	0	0	0	0
3. Can knowledge of concepts be used to answer the item?	0	0	0	0
4. Can knowledge of principles be used to answer the item?	0	0	0	0
5. Can knowledge of relationships be used to answer the item?	0	0	0	0
6. Does item have a TGF* already completed/needs to be completed?	0	1	1	0
7. Does item refer directly or indirectly to info. in a completed & separate TGF (g/s)?	0	0	0	1
8. Does item refer to info. in a tTGF* (s)* separate from stem?	0	0	0	0
9. Does item have (or refers to info. in) a completed TGF (g/s)*?	0	0	0	0
10. When present, is it possible to use info. in completed TGF (g/s) to answer item?	0	0	0	1
11. Is it necessary to use info. in completed TGF (g/s) to answer item?	0	0	0	0
12. Is some of the info. needed to answer item in TGF (s)?	0	0	0	0
13. Is all info. needed to answer item in tTGF in block with item? [All info. is (g)]	1	1	0	1
14. Is all info. needed to answer item in tTGF in block with item? [Some info. is (s)]	0	0	0	1
15. Does response require a TGF to be drawn or completed?	0	1	1	0
16. Does response require a GF to be drawn or completed?	0	0	0	0
17. Does item have science vocabulary that must be understood to answer item?	0	0	0	0
18. Must response meet all conditions specified in stem?	0	0	0	0
19. Does item have hypotheticals/exceptions/negations that make item complex?	0	0	0	0
20. Can item be solved by choosing the odd option out?	0	0	0	0
21. Does item require info. that could have been gained through practical experience?	0	0	0	0
22. Does item require only info. in item itself (incl. procedural knowledge)? [Not (s)]	0	0	0	0
23. Can reasoning from general concept/principle/law to specific conclusion be used?	0	1	0	1
24. Is tracing cause-effect from one component to another in system needed to answer?	0	1	0	1
25. Can formal inductive reasoning be used to answer item?	0	0	0	0
26. Can application of concept/principle/idea be used to answer item?	0	1	0	1
27. Can thinking with models/analogies be used to answer item?	0	0	0	1
28. Is generation of hypothesis/prediction necessary to answer item?	0	0	0	0
29. Does item require ident. of variables/controls in design of test for hypothesis?	0	0	0	0
30. Does item require generating operationalized procedures for testing a hypothesis?	0	0	0	0

31. Does item require use of multiple control groups in design of test for hypothesis?	0	0	0	0
32. Does item require that a response be given and the response be justified?	0	0	1	0
33. Is item a 4-category constructed-response item?	1	0	0	0
34. Is item a short constructed-response item?	0	1	1	0
35. Does item stem have one or more intratext referentials (e.g., it, this, these)?	0	1	1	1
36. Does item stem have one or more clauses with fronted structures?	1	0	0	1

* TGF = table, graph, or figure
tTGF = text, table, graph, or figure

(g) = given
(s) = student-generated