



Research Report

No. 2009-3

Validating Cognitive Models of Task Performance in Algebra on the SAT[®]

**Mark J. Gierl, Jacqueline P. Leighton,
Changjiang Wang, Jiawen Zhou,
Rebecca Gokiert, and Adele Tan**

Validating Cognitive Models
of Task Performance in
Algebra on the SAT®

**Mark J. Gierl, Jacqueline P. Leighton, Changjiang Wang, Jiawen Zhou,
Rebecca Gokiert, and Adele Tan**

The College Board, New York, 2009

Mark J. Gierl is a professor and Canada research chair at the University of Alberta.

Jacqueline P. Leighton is a professor and registered psychologist at the University of Alberta.

Changjiang Wang is a doctoral student at the University of Alberta.

Jiawen Zhou is a doctoral student at the University of Alberta.

Rebecca Gokiert is a doctoral student at the University of Alberta.

Adele Tan is a doctoral student at the University of Alberta.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,600 schools, colleges, universities and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,800 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities and concerns.

For further information, visit www.collegeboard.com.

Additional copies of this report (item #090482922) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800-323-7155. The price is \$15. Please include \$4 for postage and handling.

© 2009 The College Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT and Student Search Service are registered trademarks of the College Board. *inspiring minds* and SAT Reasoning Test are trademarks owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners. Visit the College Board on the Web: www.collegeboard.com.

Printed in the United States of America.

Contents

<i>Introduction</i>	1	4. Summary of the Attributes Required to Solve the Items in Hierarchy 3, Ratios and Algebra.	22
<i>Cognitive Models and Educational Measurement</i>	1	5. Summary of the Attributes Required to Solve the Items in Hierarchy 4, Equation and Inequality Solutions, Algebraic Operations, Algebraic Substitution, and Exponents	26
<i>Research Methods</i>	5	6. Summary of the Strategies Used to Correctly Solve Items but Excluded from the Cognitive Models	31
<i>Protocol Analysis</i>	5	7. Six Steps in Model-Based Reasoning	33
<i>Data Collection Steps</i>	6		
<i>Step 1: Creating the SAT® Algebra Subtest</i>	6		
<i>Step 2: Collecting the Verbal Protocol Data</i>	6		
<i>Step 3: Developing Cognitive Flowcharts and Mapping Verbal Reports onto the Attributes and Hierarchies</i>	7		
<i>Results</i>	9		
<i>Student and Item Characteristics</i>	9		
<i>Protocol Summary and Attribute Comparison</i>	9		
<i>Summary and Implications for Diagnostic Assessment Using the AHM</i>	28		
<i>Summary of Current Study</i>	28		
<i>Implications of Verbal Protocol Results for Diagnostic Assessment with the AHM</i>	29		
<i>Model Modifications</i>	29		
<i>Diagnostic Inferential Errors</i>	30		
<i>The Evolution of Cognitive Models</i>	32		
<i>References</i>	34		
<i>Tables</i>			
1. Psychometric Characteristics for the Think-Aloud and March 2005 Samples on the 21 Algebra Items	9		
2. Summary of the Attributes Required to Solve the Items in Hierarchy 1, Basic Algebra I	10		
3. Summary of the Attributes Required to Solve the Items in Hierarchy 2, Basic Algebra II	16		
4. Summary of the Attributes Required to Solve the Items in Hierarchy 3, Ratios and Algebra.	22		
5. Summary of the Attributes Required to Solve the Items in Hierarchy 4, Equation and Inequality Solutions, Algebraic Operations, Algebraic Substitution, and Exponents	26		
6. Summary of the Strategies Used to Correctly Solve Items but Excluded from the Cognitive Models	31		
7. Six Steps in Model-Based Reasoning	33		

Figures

1. Four cognitive hierarchies used to describe student performance on the SAT algebra subtest, as presented in Gierl, Wang et al. (2007)	3
2. Basic Algebra I, Hierarchy 1	10
3. Problem-solving strategies for Attribute 1 in Hierarchy 1	11
4. Problem-solving strategies for Attribute 2 in Hierarchy 1	12
5. Problem-solving strategies for Attribute 3 in Hierarchy 1	13
6. Problem-solving strategies for Attribute 4 in Hierarchy 1	14
7. Problem-solving strategies for Attribute 5 in Hierarchy 1	15
8. Basic Algebra I, Hierarchy 1 (revised)	15
9. Basic Algebra II, Hierarchy 2	16
10. Problem-solving strategies for Attribute 2 in Hierarchy 2	17
11. Problem-solving strategies for Attribute 3 in Hierarchy 2	18
12. Problem-solving strategies for Attribute 6 in Hierarchy 2	19
13. Problem-solving strategies for Attribute 7 in Hierarchy 2	19

14. Problem-solving strategies for Attribute 8 in Hierarchy 2	20
15. Problem-solving strategies for Attribute 9 in Hierarchy 2	21
16. Ratios and algebra, Hierarchy 3	21
17. Problem-solving strategies for Attribute 1 in Hierarchy 3	22
18. Problem-solving strategies for Attribute 2 in Hierarchy 3	23
19. Problem-solving strategies for Attribute 3 in Hierarchy 3	23
20. Problem-solving strategies for Attribute 4 in Hierarchy 3	24
21. Problem-solving strategies for Attribute 7 in Hierarchy 3	24
22. Problem-solving strategies for Attribute 8 in Hierarchy 3	25
23. Problem-solving strategies for Attribute 9 in Hierarchy 3	25
24. Equation and inequality solutions, algebraic operations, algebraic substitution, and exponents, Hierarchy 4	26
25. Problem-solving strategies for Attribute 3 in Hierarchy 4	27
26. Problem-solving strategies for Attribute 4 in Hierarchy 4	27

Introduction

The attribute hierarchy method (AHM) (Leighton, Gierl, and Hunka, 2004) is a psychometric procedure for classifying examinees' test item responses into a set of structured attribute patterns associated with different components from a cognitive model of task performance. An attribute is a description of the procedural or declarative knowledge needed to perform a task in a specific domain. These attributes form a hierarchy that defines the psychological ordering among the attributes required to solve test items. The examinee must possess these attributes to answer the items correctly. The attribute hierarchy serves as a cognitive model of task performance, which refers to a simplified description of human problem solving on standardized tasks that facilitates explanation and prediction of students' performance (Leighton and Gierl, 2007a). A cognitive model of task performance is specified at a small grain size because it accentuates the cognitive procedures underlying test performance. Assessments based on cognitive models of task performance should be developed so that items directly measure specific cognitive processes of increasing complexity, leading to strong inferences about examinees' cognitive skills. By using the attribute hierarchy to create items to measure the cognitive components described by the model, the test developer can orchestrate which attributes are measured by which items. By using the attribute hierarchy to interpret test performance, the test developer gains control over the scores and the inferences about processes and skills associated with test performance. Consequently, the attribute hierarchy has a foundational role in the AHM, as it represents both the construct and the cognitive skills that underlie test performance.

In a report, Gierl, Wang, and Zhou (2007) applied the AHM to a sample of algebra items administered on the SAT® in March 2005. The cognitive model of task performance used to develop the attribute hierarchy for the analyses was generated by the investigators who conducted a task analysis of the SAT Algebra I and II items to identify the mathematical concepts, operations, procedures, and strategies that students might use to solve items on the SAT. However, the evidence Gierl, Wang, et al. collected to support the algebra models was limited because it was only based on a task analysis of the items. Hence, the purpose of the current study is to present research focused on validating the four algebra cognitive models in Gierl, Wang, et al., using student response data collected with protocol analysis methods to evaluate the knowledge structures and processing skills used by a sample of SAT test-takers. The verbal protocol data were collected in November 2005 by asking 21 students to think aloud as they solved a sample of 21 algebra items taken from the March

2005 administration of the SAT. The structure of this report is as follows: In the first section we define the phrase "cognitive model" in educational measurement, and we explain why these models are important in the development and analysis of diagnostic assessments. We also provide a sample of cognitive models that can be used to characterize student performance in algebra. In the second section we describe the methods used to collect and analyze the verbal protocol data. In the third section we present the results from our protocol analysis. In the fourth section we provide a summary of our study and highlight the implications of our results for making cognitive diagnostic inferences using the AHM.

Cognitive Models and Educational Measurement

To make specific inferences about problem solving, cognitive models are required. A cognitive model in educational measurement refers to a "*simplified description of human problem solving on standardized educational tasks, which helps to characterize the knowledge and skills students at different levels of learning have acquired and to facilitate the explanation and prediction of students' performance*" (Leighton and Gierl, 2007a, p. 5). These models provide an interpretative framework that can guide item development so that test performance can be linked to specific cognitive inferences about examinees' knowledge, processes, and strategies. These models also provide the means for connecting cognitive principles with measurement practices, as Snow and Lohman (1989) explain:

As a substantive focus for cognitive psychology then, "ability," the latent trait (θ in EPM [educational and psychometric measurement] models), is not considered univocal, except as a convenient summary of amount correct regardless of how obtained. Rather, a score reflects a complex combination of processing skills, strategies, and knowledge components, both procedural and declarative and both controlled and automatic, some of which are variant and some invariant across persons, or tasks, or stages of practice, in any given sample of persons or tasks. In other samples of persons or situations, different combinations and different variants and invariants might come into play. Cognitive psychology's contribution is to analyze these complexes. (pp. 267–268)

Cognitive processes represent a sequence of internal events where information in short- and long-term memory interacts. Short-term memory is seen as a

storage system associated with limited capacity, fast access, and conscious awareness. Long-term memory is seen as a storage system associated with unlimited capacity, slow access, and unconscious awareness. Verbal reports provide a description of the examinees' thought processes when the reported information enters short-term memory. Information in long-term memory can be made available to conscious awareness if it is transferred into short-term memory. However, until this information is accessed and attended to, it will not be consciously experienced. Given these assertions about information processing, cognitive models of task performance can be generated by studying the processes used by examinees as they respond to items on tests. These models can be created by having examinees think aloud as they solve tasks in a specific domain or content area in order to identify the information requirements and processing skills elicited by the tasks (Ericsson and Simon, 1980, 1993; Leighton, 2004; Leighton and Gierl, 2007b; Royer, Cisero, and Carlo, 1993; Taylor and Dionne, 2000). The model is then evaluated by comparing its fit to the examinees' observed response data and to competing models as a way of substantiating the components and structure. After extensive evaluation, scrutiny, and revision, the model may also generalize to other groups of examinees and different problem-solving tasks.

A cognitive model of task performance is specified at a small grain size to magnify the cognitive processes underlying test performance. Often, a cognitive model of task performance will also reflect a hierarchy of cognitive processes within a domain because cognitive processes share dependencies and function within a much larger network of interrelated processes, competencies, and skills (Anderson, 1996; Dawson, 1998; Fodor, 1983; Kuhn, 2001; Mislevy, Steinberg, and Almond, 2003). Assessments based on cognitive models of task performance should be developed so that test items directly measure specific cognitive processes of increasing complexity in the understanding of a domain. The items can be designed with this hierarchical order in mind, so that test performance is directly linked to information about students' cognitive strengths and weaknesses. Strong inferences about examinees' cognitive skills can be made because the small grain size in these models helps illuminate the knowledge and skills required to perform competently on testing tasks. Specific diagnostic inferences can also be generated when items are developed to measure different components and processes in the model.

The strength of developing test items according to a cognitive model of task performance stems from the detailed information that can be obtained about the

knowledge structures and processing skills that produce a test score. Each item is designed to yield specific information about students' cognitive strengths and weaknesses. If the target of inference is information about students' cognitive skills, then the small grain size associated with these models is required for generating specific information. This specific information can be generated because the grain size of these models is narrow, thereby increasing the depth to which both knowledge and skills are measured with the test items. A cognitive model of task performance also requires empirical support with psychological evidence from the populations to which inferences will be targeted. Once this model is validated with the population of interest, items can be created that measure specific components of the model, thereby providing developers a way of controlling the specific cognitive attributes measured by the test.

The challenges inherent to developing items according to a cognitive model of task performance stems from the paucity of information currently available on the knowledge, processes, and strategies that characterize student performance in most testing situations. Because little is known about how students actually solve items on educational tests, relatively few models exist. Even when these models are available, they rarely guide psychometric analyses because they are usually restricted to a narrow domain; they are expensive to develop initially and to refine over time because they require extensive—typically experimental—studies of problem solving on specific tasks; and they require cognitive measurement expertise, which is uncommon.

To illustrate an application of cognitive model development, as it can be applied to the AHM within the domain of mathematics, Gierl, Wang et al. (2007) developed four cognitive hierarchies to account for examinee performance in algebra. The algebra hierarchies are based on a task analysis of the released items from the March 2005 administration of the SAT. Sample algebra items from the SAT mathematics section can be accessed through the College Board Web site at www.collegeboard.com.

The SAT is a standardized test designed to measure college readiness. Both critical thinking and reasoning skills are evaluated. The mathematics section contains items in several content areas: number and operations; Algebra I, II, and functions; geometry; and statistics, probability, and data analysis. Multiple-choice and constructed-response item formats are used, but the items for both formats are scored dichotomously. All items in Algebra I and II were evaluated and used to develop the algebra hierarchies.

As previously noted, cognitive models of task performance guide diagnostic inferences because they are specified at a small grain size and they magnify the cognitive processes that underlie performance. Unfortunately, few cognitive models currently exist. Ideally, a theory of task performance would direct the development of a cognitive model. But in the absence of such a theory, a cognitive model must still be specified to create the attribute hierarchy. Another starting point is to create a model from a task analysis conducted within the domain of interest. In conducting the task analysis of the SAT algebra items, Gierl, Wang et al. (2007) first solved each test item and attempted to identify the mathematical concepts, operations, procedures, and strategies used to solve each item. They then categorized these cognitive attributes so they could be

ordered in a logical, hierarchical sequence to summarize problem-solving performance. Four plausible cognitive models of algebra performance were identified, as presented in Figure 1. Each one of these models could be used to characterize performance on a subset of items administered in the Algebra I and II sections of the SAT. *That is, each model serves as a hypothesized structure for describing the cognitive skills required to solve items in algebra.* Our task in the current study is to evaluate these hypotheses. The attributes are labeled A1 to A η , where η is the number of attributes. The test items identified from the March 2005 SAT administration that measured the attributes are labeled at the right side of each attribute. The items are labeled 1 to 21. The complete set of algebra items is available from the College Board or from the first author.

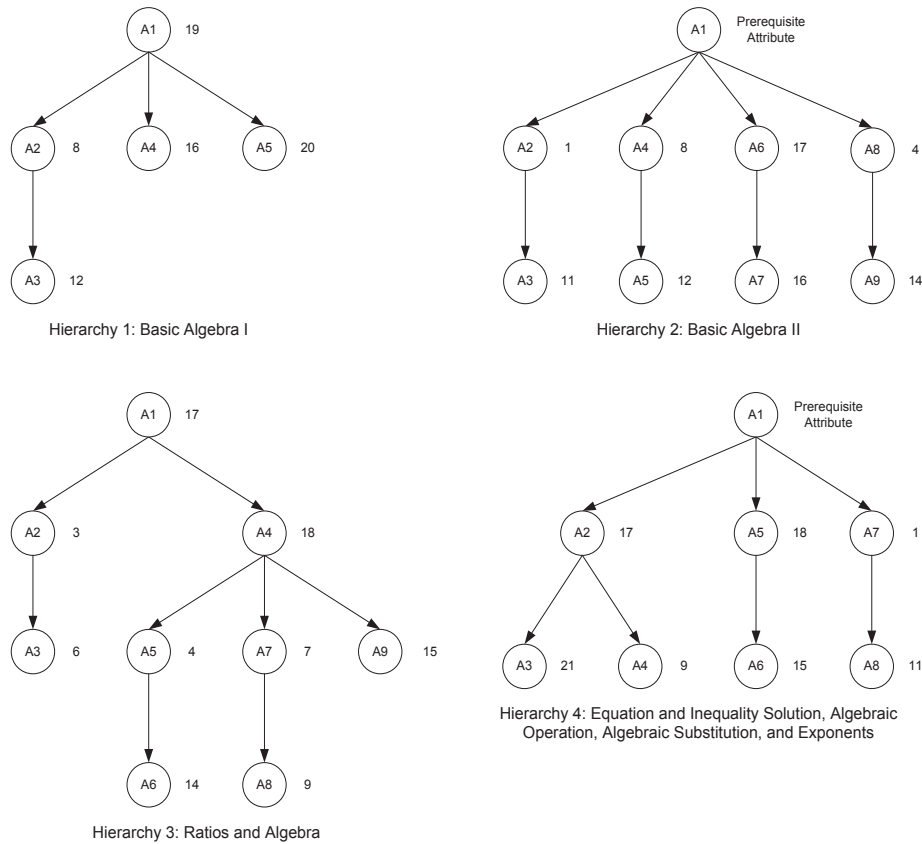


Figure 1. Four cognitive hierarchies used to describe student performance on the SAT algebra subtest, as presented in Gierl, Wang et al. (2007). These four models guided the AHM analyses presented in Gierl, Wang et al. using a random sample of 5,000 students who took the items during the March 2005 administration.

Research Methods

Protocol Analysis

One method used to gain information about the representations and processes in cognition is to probe the students' internal states from their overt verbal responses using protocol analysis (Ericsson and Simon, 1980, 1993). Verbal think-aloud protocols provide one source of evidence that a student has reached a solution to a problem. It also provides a method for tracing and documenting the representations and processes used by students to generate a solution. Two types of think-aloud reports are used in protocol analysis: concurrent and retrospective verbalization (Ericsson and Simon, 1980, 1993). These two types of reports yield different information about the students' cognitive procedures.

Concurrent verbalizations yield information at the time the student is attending to the information. Concurrent verbalizations are elicited by asking students to think aloud as they solve test problems. Concurrent verbalizations can take the form of "talk aloud" or type 1 verbalizations, where various kinds of information in short-term memory are reported by the student at the time they are attended to. Concurrent verbalizations can also be "think aloud" or type 2 verbalizations, where one or more mediating processes are believed to have occurred before the information is verbalized by the student. Type 1 verbalizations are commonly reported for elementary tasks such as simple pattern recognition, while type 2 verbalizations are elicited by more complex tasks such as algebra problem solving.

Retrospective verbalizations are obtained by asking students to "report everything you can remember about your thoughts during the last problem" immediately after the problem is completed. Retrospective verbalizations, or type 3 verbalizations, are verbal reports about cognitive processes that occurred at an earlier point in time. They contain some information that is retrieved from long-term memory, and therefore require the student to make inferences about the cognitive processes used for the task. Concurrent verbalization, by comparison, yields relatively direct access to short-term memory and does not require inferential steps.

The dominant theory for protocol analysis was developed by Ericsson and Simon (1980, 1993). They used the theoretical framework of human information processing (IP) to describe a model for obtaining and interpreting verbal report data. Ericsson and Simon specify how the IP system operates and how verbal report data are produced. Their theory begins with a key assumption: Short-term memory has a limited storage capacity; therefore, only the most recently heeded or attended-to information is accessible. Consequently, concurrent verbalizations or think-aloud reports can be used to access this information. Ericsson and Simon also

assume that a portion of the information in short-term memory is retained in long-term memory before it is lost. This portion of retained information can be retrieved from long-term memory at a later time. As a result, retrospective verbalizations can be used to access the retained information immediately (i.e., within 10 seconds) after problem solving is completed (Ericsson and Simon, 1993, p. xvi).

The level of information processing obtained from verbal reports is well defined, according to Ericsson and Simon. The information in short-term memory contains cognitive representations and processes that only go down to a modest level of detail, although the specific details would depend on the specific strategies used by students and the nature of the information stored in long-term memory. Ericsson and Simon (1980) explain: "We would not expect to find information about simple, automated processes, much less neuronal events. Thus, the architecture of the control apparatus determines the fineness of grain of the representations and processes in short-term memory" (p. 225). In other words, the granularity of the cognitive process data is determined, for the most part, by the data collection method. Ericsson and Simon also argue that the verbal encoding processes involved in thinking aloud do not change the structure or the nature of the information in short-term memory, although they may decrease the speed of problem solving. For the most part, these assumptions have been scrutinized and empirically validated, as there is considerable evidence to support their model.

Verbal reports have been used to validate educational tests, although the number of published studies using this method is relatively small. Norris (1990), for example, used verbal reports to validate a multiple-choice test in critical thinking as well as to evaluate the impact of concurrent and retrospective verbalizations on test performance. Norris found no test score differences across four verbal reporting conditions when they were compared to a paper-and-pencil condition with no verbal reporting. He concluded:

The verbal reports of thinking collected for this study contained a wealth of information useful for rating the quality of subjects' thinking and for diagnosing specific problems with items. . . . Given the results of this study, it is reasonable to trust the diagnostic information as an accurate representation of problems that would occur with the items taken in a paper-and-pencil format. Moreover, the verbal reports provide more direct information on the exact nature of the problems of these sorts than that provided by traditional item analysis statistics. (p. 55)

The results from the Norris study strongly suggest that verbal reports on multiple-choice tests do not alter either the thinking or the test performance of examinees.

To conclude, verbal think-aloud reports provide the researcher with information that students typically use to solve problems. Protocol analysis can provide insights

into cognitive problem-solving behaviors that map onto the strategies students use when solving similar problems without verbal reports. The procedures used to elicit verbal reports are extremely important because they may reflect different cognitive processes. Nevertheless, it is safe to conclude, as Ericsson and Simon (1980, p. 247) do, that “verbal reports, elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and thoroughly reliable source of information about cognitive processes” (see also Leighton, 2004).

Data Collection Steps

This study was conducted in three stages. First, an algebra subtest using 21 algebra items taken from the March 2005 administration of the SAT was created. Second, the algebra subtest was administered in November 2005 to 21 high school students from New York City. Students were asked to think aloud as they solved items on the algebra subtest and, after they solved the problem, to report how they arrived at their answers. These testing sessions were not timed, as a regular administration of the SAT would be. Third, flowcharts were created to represent students’ cognitive processes as reported in the think-aloud protocols. These flowcharts were used to evaluate both the item attributes and their hierarchical ordering for the algebra subtest, as shown in Figure 1.

Step 1: Creating the SAT® Algebra Subtest

The March 2005 administration of the SAT had 21 items in Algebra I and II. Initially, we selected items for inclusion using two criteria. First, we wanted to include items with a range of difficulty levels. Hence, the items were ordered from least to most difficult. Second, we wanted items with a range of College Board skill codes, as identified by O’Callaghan, Morley, and Schwartz (2004). These skills included applying basic mathematics knowledge (Skill 1), applying more advanced mathematics knowledge (Skill 2), managing complexity (Skill 3), and modeling and insight (Skill 4). Taken together, three blocks of six items were constructed, where each block would have a range of difficulty levels and include each skill code at least once in the block. From these rules, our initial set of items was selected. Because each session lasted 60 minutes, our goal was to collect complete data from all participants on the 18 items. But we also included the remaining three items in each test booklet. These three items would be administered, if time permitted. These three items were of moderate to high difficulty and measured only Skills 3 and 4. All students were able to complete the 21 items. As a result, our protocol analyses were conducted on all 21 Algebra I and II items from the March 2005 administration of the SAT.

Step 2: Collecting the Verbal Protocol Data

The algebra subtest was administered in November 2005 to 21 students (12 males, 9 females) who attended school in New York City. The sample was drawn from all potential New York City students who took the PSAT/NMSQT® as tenth-graders, with the following six constraints: (1) the assessment was administered without accommodations; (2) students live and attend school in New York City; (3) students scored between 550 and 650 on mathematics; (4) students scored between 600 and 800 on critical reading; (5) students opted in to the Student Search Service®; and (6) students had only taken the PSAT/NMSQT once. A statistical analyst at the College Board then sampled from this population, producing a list of 75 male and 75 female students who were eligible to participate. All 150 students were contacted by mail. Of the 150 students contacted, 26 agreed to participate (17.3 percent of the total sample); of the 26 who agreed, 21 students attended their scheduled testing session at the College Board main office. Each student who participated received \$50 and a public transportation voucher for travel to and from the College Board. Student testing started on November 11, 2006, and ended on November 13, 2006.

Each volunteer was individually assessed in an empty conference room at the College Board’s main office. Students were asked to think aloud as they solved the items. Each session was audiotaped, and the testing sessions were not timed. Students received the following instructions:

Thank you for agreeing to participate in today’s study. Please know that your participation is completely voluntary and you are free to go at any time. Now, let me explain what we will be doing today for about 60 minutes.

In this study we are interested in what goes through your mind or what you think about when you find answers to SAT questions in math. In order to do this I’m going to ask you to THINK ALOUD as you work on the problems given. What I mean by think aloud is that I want you to tell me EVERYTHING you are thinking from the time you first see the question until you give an answer.

I would like you to talk aloud CONSTANTLY from the time I present each problem until you have given your final answer to the question. I don’t want you to try to plan out what you say or try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself. It is most important that you keep talking. If you are silent for any long period of time I will remind you to talk. Do you understand what I want you to do?

I will tape record our session because I want to get an accurate record of your think-aloud reports. Please know that all the information you share today with me

will be kept confidential and anonymous. Do you have any questions?

Students were given two practice items before each session began. Students were debriefed at the end of the session, where they received a complete description of the study and its purpose.

After students reported an answer for the algebra item, they were then asked, “How did you figure out the answer to the problem?” unless the student volunteered the information. If the students’ description was unclear, follow-up questions were asked. For example, if a student said, “I just remembered,” he or she was asked, “What did you remember?” Because the focus in this study was to understand how students solved problems rather than why students solved problems the way they did, additional probes were not used. Students were not given feedback on their performance during the test. Each session lasted 45 minutes on average.

Step 3: Developing Cognitive Flowcharts and Mapping Verbal Reports onto the Attributes and Hierarchies

Flowcharts were created to represent students’ cognitive procedures, as reported in the think-aloud protocols.

1. **Start/Stop Box**—This is a parabola that starts and stops the flowchart. In this study students began by reading the questions out loud. Therefore, the start box represents this point in the problem-solving sequence. The protocol was complete when students reported their answer. Thus, the stop box contained the students’ final answer. Only the solution used to reach the final answer was graphed and presented in this study.
2. **Process Box**—This is a rectangle with one flow line leading into it and one leading out of it. Each process box contained an elementary cognitive process reported by the students as they solved the items.
3. **Connector**—This is a circle connecting two flow lines in a diagram. In most cases, connectors represented junctions or links in the flowchart where students differed from one another.
4. **Flow Line**—This is a line with a one-way arrow used to connect process boxes with one another or with process boxes that have start/stop boxes. Flow lines indicated the direction of information processing as students worked toward their solutions. Information was assumed to flow as a sequential rather than parallel process; therefore, only one elementary event is processed at a time and only one arrow per box is presented.

In the second stage, the elementary cognitive processes in the flowcharts were coded into more general categories associated with specific problem-solving strategies. For example, in a problem such as “ $4(x - 1) - 3x = 12$, then $x = ?$,” students typically use as many as five different elementary processes. However, these processes were indicative of a more general problem-solving strategy—namely, solve x by isolating the variable on one side of the equation. Both the elementary cognitive processes and the problem-solving strategies used by students to solve each of the 21 SAT algebra items were graphed. Although both correct and incorrect responses were coded, only the correct responses are presented in this report. The decision to focus on correct responses stems from the nature of our psychometric procedure, the AHM, which is used to model

Flowcharts were needed to condense the qualitative data produced by the think-aloud procedure, as 21 audiotapes with more than 16 hours of verbal reports were collected during Step 2. The cognitive flowcharts were created and coded in three stages. In the first stage, two graduate assistants on this project (Wang and Zhou) listened to each audiotape and created a flowchart for each student protocol. The elementary cognitive process reported by students for each item was documented and graphed using both the students’ verbal responses and their written responses. Flowcharts were used because they provided a systematic method for representing problem solving where both the components (i.e., elementary cognitive processes) and the overall structure of the components could be graphically presented. Flowcharts also highlighted individual differences where the elementary steps and solution strategies for each student could be compared and contrasted. Standard flowchart symbols, as found in cognitive and computer science, were followed (see, for example, Reddy and Ziegler, 1994). The flowcharts contained four different symbols:

correct response patterns. While incorrect responses can be a valuable source of diagnostic information (Luecht, 2007), these data cannot currently be modeled with the AHM.

Finally, in the third stage, to evaluate how well the attributes and the hierarchies specified in the Figure 1 cognitive models developed by Gierl, Wang et al. (2007) matched the cognitive processes reported by students, the attribute descriptions were compared to the cognitive flowcharts for each item. Two raters (Wang and Zhou) were asked to independently compare the student think-aloud flowchart data to those of the cognitive models. Once the comparison was complete, the reviewers met to discuss their results with one another and with the first author of this study. All disagreements were discussed, debated, and resolved.

Results

The results are presented in two sections. We begin by describing the student sample and the item characteristics for both the 21-student sample as well as for a random sample of 5,000 students who answered these 21 algebra items on the March 2005 administration. We then present a summary of the protocol analysis, which highlights the similarities and differences between the cognitive models described in Gierl, Wang et al. (2007) relative to the student verbal report data.

Student and Item Characteristics

The first set of descriptive analyses was conducted to assess the student characteristics in our sample. Twenty-one students (12 males and 9 females) were included. The mean PSAT/NMSQT critical reading score was 65 (standard deviation [SD]=4.21); the mean PSAT/NMSQT mathematics score was 60 (SD=3.64). Hence, our sample had a higher critical reading and mathematics score than the 5,000-student sample, as was our goal from the sampling plan. We intentionally selected students with above-average PSAT/NMSQT critical reading scores, as these students were expected to have stronger verbal skills and thus be more proficient at verbalizing their thinking processes. At the same time, we attempted to select students with average to above-average math skills so that

a range of mathematical proficiencies would be included in the think-aloud sample. These selection decisions may limit the generalizability of our results, but it did help ensure that the verbal reports were clearly articulated and therefore easier to code. Sixteen of the 21 students were white, 1 student was an Asian/Pacific Islander, 1 student was black/African American, and 1 student was Hispanic. Two students did not respond to the ethnicity self-report item.

The second set of descriptive analyses was conducted to evaluate the item characteristics of the 21-item SAT algebra subtest, and to compare these results to those of the sample of 5,000 students who answered the same items in March 2005. Using data from the 21 students, the mean performance on the 21-item algebra subtest was 16.48 (SD=1.81), the mean item difficulty value was 0.78 (SD=0.23), and the mean item discrimination value was 0.39 (SD=0.34). Because the think-aloud sample was selected to have a higher and more restricted SAT mathematics score, the item characteristics were lower for the random sample of 5,000 students who took the March 2005 administration of the SAT. Using data from the 5,000-student sample, the mean performance on the 21-item algebra subtest was 12.11 (SD=4.56), the mean item difficulty value was 0.58 (SD=0.24), and the mean item discrimination value was 0.68 (SD=0.09). The results are summarized in Table 1.

Table 1

Psychometric Characteristics for the Think-Aloud and March 2005 Samples on the 21 Algebra Items

<i>Sample</i>	<i>Think-Aloud</i>	<i>March 2005</i>
No. of Examinees	21	5,000
No. of Items	21	21
Mean	16.48	12.11
SD	1.81	4.56
Mean Item Difficulty	0.78	0.58
SD Item Difficulty	0.23	0.24
Mean Item Discrimination ^a	0.39	0.68
SD Item Discrimination	0.34	0.09

^aBiserial correlation

Protocol Summary and Attribute Comparison

In the next section the protocol analyses are presented with the following information: (a) the initial cognitive model of task performance from Gierl, Wang et al. (2007) (each model includes the *p*-values calculated from the sample of 5,000 students who answered the items in March 2005, the College Board ability-level classification,

and the College Board skill code from O'Callaghan et al. [2004]); (b) a description of the attribute in each model from Gierl, Wang et al.; (c) the Hierarchy Classification Index [HCI_i] for assessing model-data fit (recall that this index ranges from +1, indicating that the examinee's observed response pattern matches the hierarchy, to -1, indicating that the examinee's observed response pattern fails to match the hierarchy—thus, higher values indicated better model-data fit [Gierl, Cui, and Hunka, 2006]);

(d) a comparison between the results from the student flowcharts with each attribute description; and, where appropriate, (e) examples from the cognitive flowcharts. Each solution path is labeled in the stop box at the bottom of the flowchart. The first solution path, for example, is

labeled “SP1.” Only those solutions leading to the correct answer are presented. Males are assigned numbers from 1 through 12, whereas females are assigned letters from A through I.

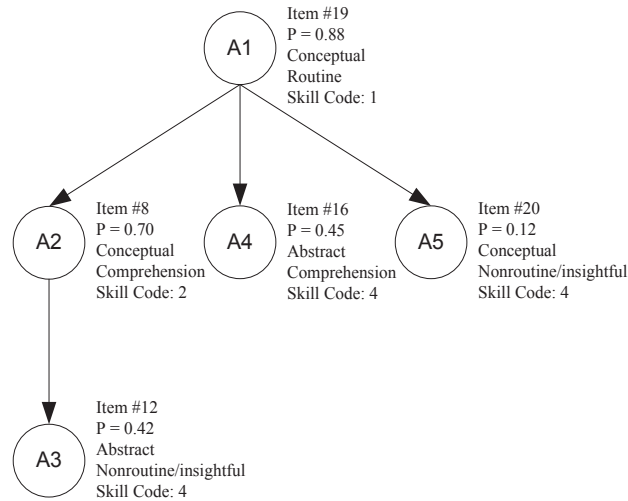


Figure 2. Basic Algebra I, Hierarchy 1.

Hierarchy 1 illustrates a cognitive model of task performance for general skills in the area of Basic Algebra I, as shown in Figure 2. It consists of five attributes: single ratio setup, conceptual geometric series, abstract geometric series, quadratic equation, and fraction transformation. This hierarchy has three independent

branches: Attributes A1, A2, and A3; Attributes A1 and A4; and Attributes A1 and A5. The HCI, for this model in Gierl, Wang et al. (2007) was 0.78, indicating relatively strong model–data fit. The complete set of attributes used in Hierarchy 1 is summarized in Table 2.

Table 2

Summary of the Attributes Required to Solve the Items in Hierarchy 1, Basic Algebra I

Attribute A1	includes the basic mathematical knowledge and skills required for setting up a single ratio by comparing two quantities.
Attribute A2	requires the mastery of the skills to order a geometric series. This attribute involves the knowledge about geometric series (e.g., the nature of the between-term ratio) and/or the consecutive numerical computation (e.g., multiplication and division).
Attribute A3	considers the skills for solving geometric series in an abstract pattern.
Attribute A4	includes the skills required for representing and executing multiple basic algebraic skills.
Attribute A5	termed fraction transformation, is also an attribute with multiple skills. This attribute requires a host of specific skills, including representing and executing multiple advanced algebraic skills, such as setting up a single ratio, skills for transforming fractions, and analysis skills such as when, where, and/or how to do the transformation.

As a prerequisite skill, Attribute A1 includes “the basic mathematical knowledge and skills required for setting up a single ratio by comparing two quantities.” Item 19 measures this attribute. For this item, students are asked to set up a single ratio to get average speed, given distance and time. All 21 students in our study correctly answered the item. Two strategies were used: retrieving the relevant knowledge about setting up a single ratio and plugging in

numbers (see Figure 3). Nineteen out of the 21 students adopted the first strategy, while the remaining 2 students adopted the second strategy. Of the 19 students who used the first strategy, two solution paths were generated. Eight students obtained the answer by directly setting up the distance/time ratio (SP1), and 11 students generated the answer by retrieving the average speed formula (SP2).

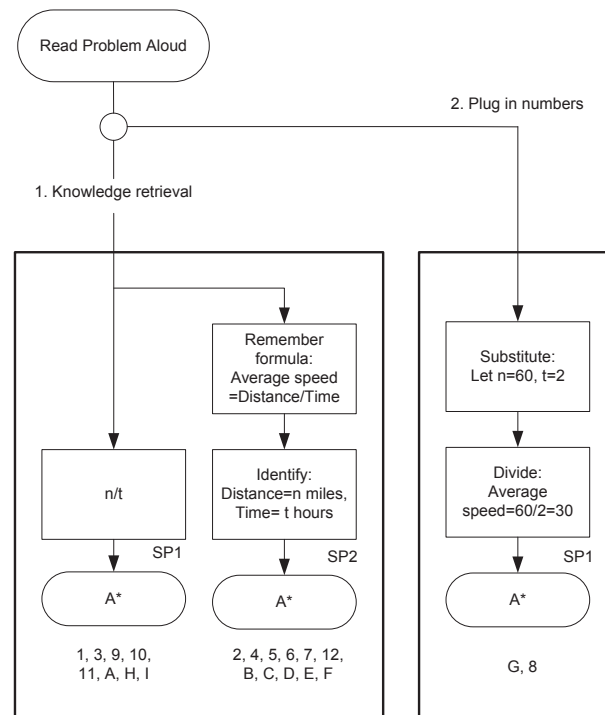


Figure 3. Problem-solving strategies for Attribute 1 in Hierarchy 1.

Attributes A2 and A3 represent the skills required to understand and solve geometric series. In addition to the knowledge and skills required in A1, A2 requires the mastery of the skills to order a geometric series. This attribute involves “knowledge about geometric series” (e.g., the nature of the between-term ratio) “and/or the consecutive numerical computation” (e.g., multiplication and division). For instance, in Item 8, students need to consecutively divide a number by the between-term ratio to get the value of a certain term in the series. Sixteen students correctly answered this item. In solving

the item, two strategies were used: the consecutive numerical computation and the use of a generic formula ($a_n = a_1 \times r^{n-1}$) (see Figure 4 on next page). Of the 16 students, 15 used the first strategy, while the remaining student used the second strategy. Of the 15 participants who used the first strategy, two solution paths were generated. Three students produced the answer by setting up and solving an equation (SP1), and 12 students produced the answer by consecutively dividing the between-term ratio (SP2).

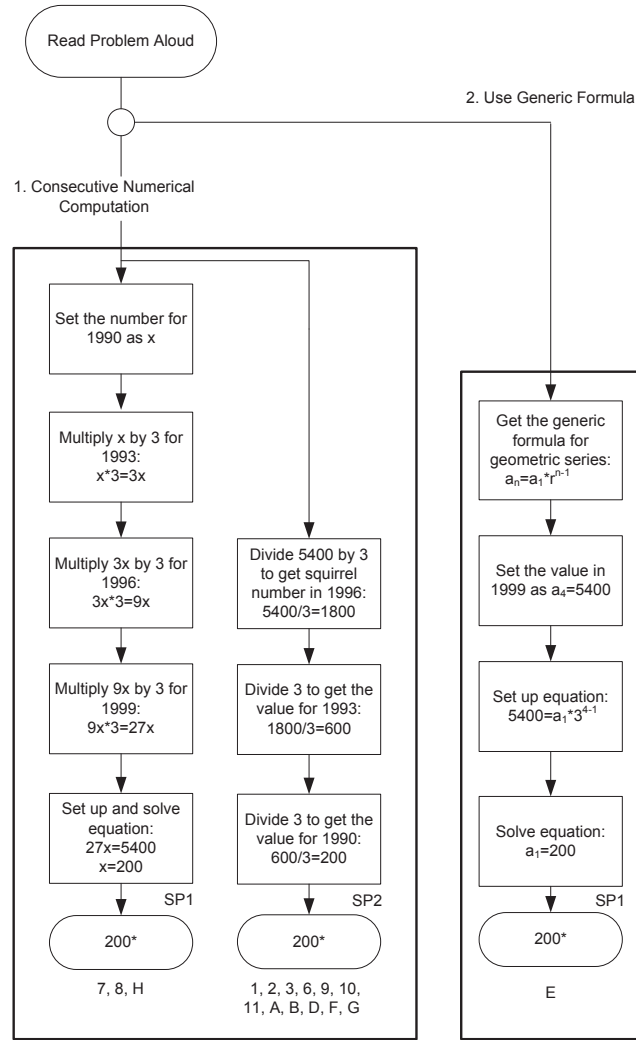


Figure 4. Problem-solving strategies for Attribute 2 in Hierarchy 1.

Attribute A3 considers the skills of solving geometric series in an abstract pattern. This attribute is an extension of Attribute A2, and in turn, Attribute A1. In addition to understanding a geometric series conceptually (i.e., Attribute A2), the skill to generate and manipulate a qualifying geometric series is required in Attribute A3. For example, in Item 12, students must compare the eighth and fifth terms in a series such as $x, 2x, 4x, \dots, 128x$, meaning they must identify the relationship between the first two terms and then set up and compute a ratio in a qualifying series to reach their solution. Eighteen students correctly answered the item. One dominant strategy was adopted by all 18 participants: namely, get the between-term ratio, write out a qualifying series with a sufficient number of terms, and divide the two terms specified in the item (see Figure 5 on next page). Depending on the type of series students wrote, two solution paths were generated. Thirteen students wrote their series in

numbers (SP1), while the remaining 5 students wrote their series in an algebraic expression (SP2).

Attribute A4, labeled “representing and executing multiple basic algebraic skills,” is another extension of Attribute A1. The multiple skills required in this attribute can be either dependent or independent. For Attribute A4, in addition to the basic skills involved in setting up a single ratio, the mastery of solving a quadratic equation is also required. This skill requires a number of simultaneous steps such as the ability to apply square, square root, multiplication, and division operators. In Item 16, to solve $\sqrt{x} = \frac{x}{40}$, students need to first square both sides of the equation to get $x = \frac{x^2}{1600}$, and then

multiply 1600 on both sides of the equation to get $1600x = x^2$. The last step can be achieved by dividing x on

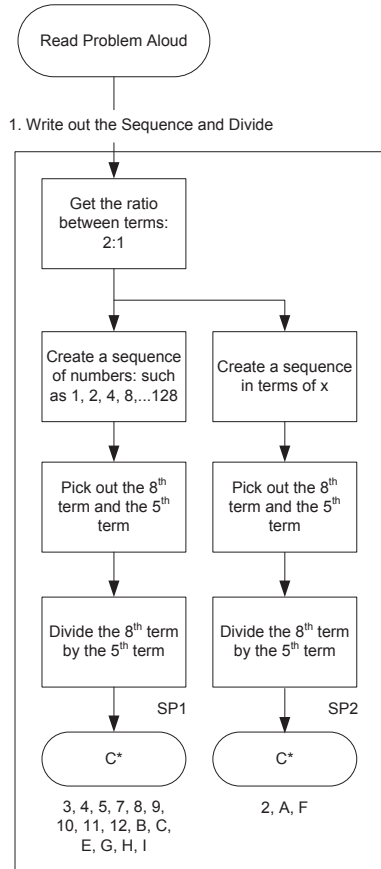


Figure 5. Problem-solving strategies for Attribute 3 in Hierarchy 1.

both sides to produce $x = 1600$. Twelve of the 21 students in our study correctly answered Item 16 (see Figure 6 on next page). Of these 12 students, 9 adopted the strategy of developing a quadratic equation solution. Depending on how the students solved the equation, three different solution paths were identified. One student solved the equation by multiplying x^{-1} on both sides of the equation (SP1); 7 students, by squaring both sides of the equation (SP2); and 1, by generating the answer directly from the equation (SP3). The remaining 3 students adopted the strategy of plugging numbers into the constructed-response item. It is important to note that although these 3 students produced the correct answer by plugging in numbers, their strategy was unrelated to our inferences about their mastery of Attribute A4, “representing and executing multiple basic algebraic skills.” Hence, this strategy, which leads to the correct answer, is inconsistent with our attribute description. This type of inconsistency is a source of error when making diagnostic inferences

with the AHM because students used skills to solve the item that are not consistent with the attribute description or the cognitive model, yet nonetheless they produce the correct answer. These types of inferential errors will be highlighted throughout our report.

Attribute A5, termed “fraction transformation,” is also an attribute with multiple skills. This attribute requires a host of specific skills “including representing and executing multiple advanced algebraic skills such as setting up a single ratio, skills for transforming fractions, and analysis skills, such as when, where, and/or how to do the transformation.” For instance, in Item 20, examinees

need to transform 1 to $\frac{y}{y}$, and add $\frac{y}{y}$ to $\frac{555 - x}{y}$ to get $\frac{555 - x + y}{y}$ in order to reach the correct answer. Six out

of 21 students correctly answered Item 20. Three strategies were used: setting up and solving the equation, plugging

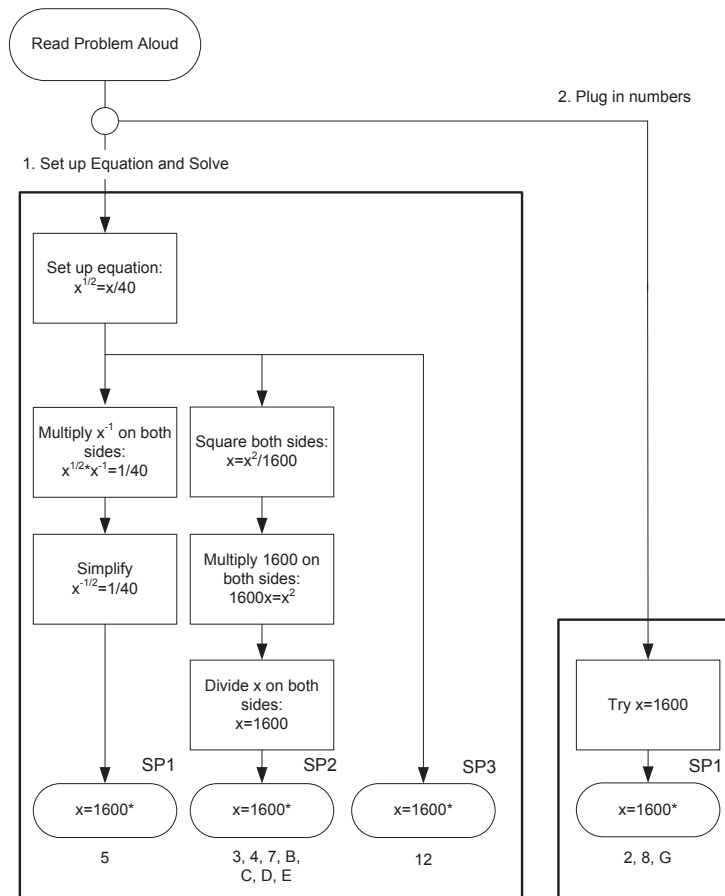


Figure 6. Problem-solving strategies for Attribute 4 in Hierarchy 1.

numbers in from the multiple-choice options, and guessing (see Figure 7 on next page). Each strategy was used by two students. However, only one of these three strategies leading to the correct solution is consistent with our attribute description: Plugging numbers in from the multiple-choice options and guessing do not measure skills associated with “representing and executing multiple advanced algebraic skills.” Therefore, these strategies, even when they produce the correct solution, serve as a source of error when making diagnostic inferences with the AHM.

Attributes A4 and A5 were viewed as hierarchically independent in the Gierl, Wang et al. (2007) task analysis. However, the results from the student protocol analysis

revealed that Attribute A5 required higher-level cognitive skills than those involved in Attribute A4 because for A5 one has to conduct an analysis of the conditions given in the item, while for A4 one only needs to apply the basic algebraic skills. The HCI_i index for the revised hierarchy was slightly less, at 0.76, than that for the initial model. However, the HCI_i outcomes are still considered comparable and within the acceptable model–data fit range. Therefore, based on the results from the protocol analysis and as supported by the HCI_i index, Hierarchy 1 was modified to include a hierarchical order between Attributes A4 and A5. The new model is shown in Figure 8 on the next page.

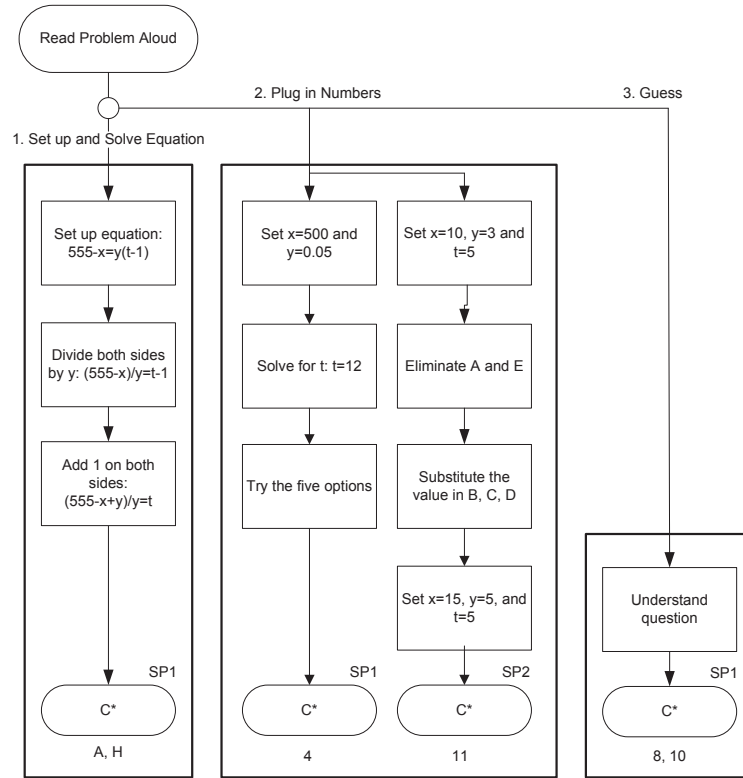


Figure 7. Problem-solving strategies for Attribute 5 in Hierarchy 1.

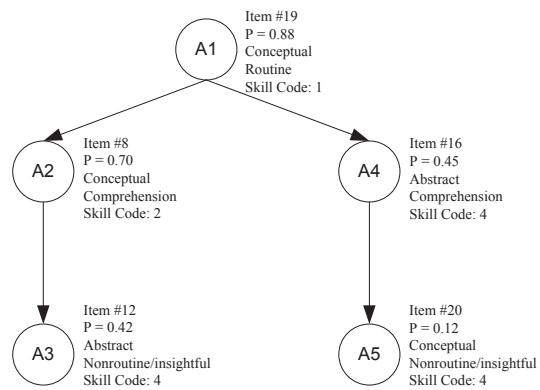


Figure 8. Basic Algebra I, Hierarchy 1 (revised).

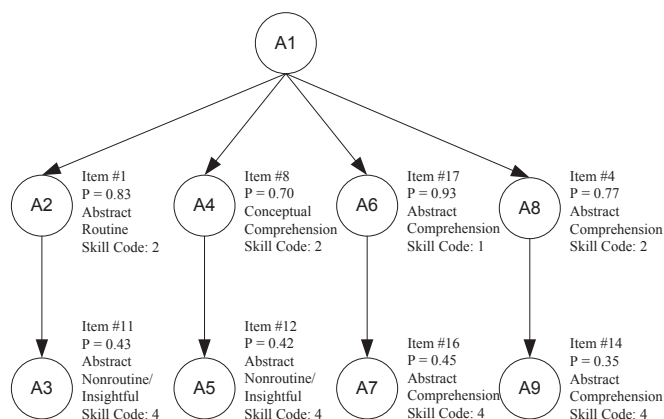


Figure 9. Basic Algebra II, Hierarchy 2.

Hierarchy 2 serves as a cognitive model of task performance for general skills in the content area of Basic Algebra II, covering the contents in exponents, geometric series, equation solution, and function graph reading. The hierarchy has four branches: Attributes A1, A2, and A3; Attributes A1, A4, and A5; Attributes A1, A6, and A7; and Attributes A1, A8, and A9. These four branches are independent from one another except that they all require

the prerequisite Attribute A1, which “includes basic language knowledge enabling students to understand the test item as well as basic mathematical knowledge and skills, such as understanding the property of absolute value and diverse but simple arithmetic operations.” The HCI for this model in Gierl, Wang et al. (2007) was 0.80, indicating strong model–data fit. The complete set of attributes used in Hierarchy 2 is summarized in Table 3.

Table 3

Summary of the Attributes Required to Solve the Items in Hierarchy 2, Basic Algebra II

Attribute A1 includes the basic language knowledge enabling students to understand the test item and basic mathematical knowledge and skills, such as the property of absolute value and arithmetic operations.

Attribute A2 includes the basic knowledge of exponential and power addition operations.

Attribute A3 involves the knowledge of power multiplication and flexible application of multiple rules in exponential operations.

Attribute A4 requires skills for ordering geometric series. This attribute involves the knowledge about geometric series (e.g., the nature of the between-term ratio) and/or the consecutive numerical computation (e.g., multiplication and division)—see also Hierarchy 1, Attribute A2.

Attribute A5 considers the skills for solving geometric series in an abstract pattern—see also Hierarchy 1, Attribute A3.

Attribute A6 requires the basic mathematical skills in solving for a linear equation (e.g., subtraction or division on both sides).

Attribute A7 requires the skills of setting up and solving for a quadratic equation, which generally involve the skills in solving a linear equation and additional skills (e.g., factoring).

Attribute A8 represents the skills of mapping a graph of a familiar function (e.g., a parabola) with its corresponding function. This attribute involves the knowledge about the graph of a familiar function and/or substituting points in the graph.

Attribute A9 deals with the abstract properties of functions, such as recognizing the graphical representation of the relationship between independent and dependent variables.

The first branch deals with basic exponential operations. Attribute A2 includes “the basic knowledge of exponential and power addition operations.” For example, in Item 1, which measures Attribute A2, the student must, above all, be able to conduct the operation that $7^n \times 7^3 = 7^{n+3}$ to correctly solve the item. Twenty out of 21 students in our study correctly answered this item. Two strategies were adopted: setting up a new equation to solve the problem, and solving the exponential equation provided in the item (see Figure 10). Seventeen out of the 20 participants used the first strategy, while the remaining 3 used the second strategy. Of the 3 students who used the second strategy, two solution paths were generated. Two students obtained the answer by directly solving the equation provided in the item (SP1), and 1 student solved the provided equation with the aid of a calculator (SP2). Clearly, the strategies used in SP2 do not reflect the skills involved in Attribute A2, “knowledge of exponential and power addition operations.” Hence, this solution path, while leading to the correct answer, is a source of error for our diagnostic inferences.

In addition to the skills in Attribute A2, Attribute A3 involves “the knowledge of power multiplication and flexible application of multiple rules in an exponential operation.” Item 11 measures Attribute A3. For this item, students are required to conduct a combination of power operations such as power multiplication and subtraction to correctly solve the item. Nineteen students correctly answered Item 11. In solving this item, two strategies were adopted: exponential computation and plugging in numbers (see Figure 11 on next page). Of the 19 students who correctly answered this item, 12 adopted the first strategy, while the remaining 7 used the second strategy. The second strategy—plugging in numbers—does not reflect the skills associated with Attribute A3, “the attribute of power multiplication and flexible application of multiple rules in exponential operation.”

The second branch in Hierarchy 2 deals with geometric series. Attributes A4 and A5, which represent “the skills required to understand and solve geometric series,” were illustrated in Hierarchy 1.

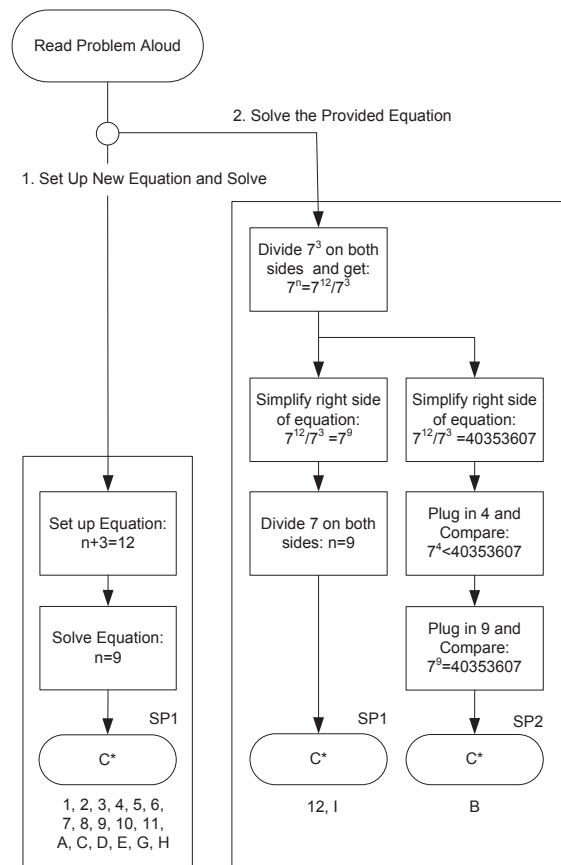


Figure 10. Problem-solving strategies for Attribute 2 in Hierarchy 2.

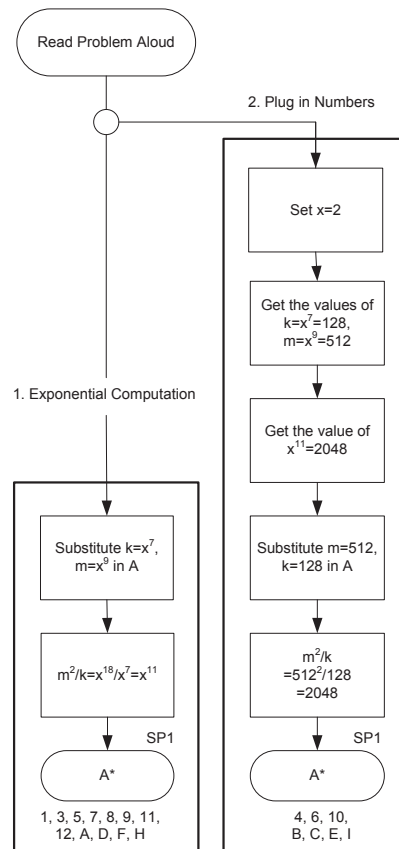


Figure 11. Problem-solving strategies for Attribute 3 in Hierarchy 2.

The third branch deals with skills involved in equation solutions. Attribute A6 requires “the basic mathematical skills in solving for a linear equation” (e.g., subtraction or division on both sides of the equation). In addition to the knowledge and skills required in A1, this attribute requires “the management of the basic mathematical skills in A1 on both sides of a linear equation.” For example, in Item 17, the examinee must treat $t + u$ as an unknown and solve the linear equation. All 21 students in our study correctly answered the item. Two strategies were involved: solving a linear equation and trying the answer options (see Figure 12 on next page). Nineteen of the 21 students adopted the first strategy, while the remaining 2 students adopted the second strategy. Of the 19 students who used the first strategy, three solution paths were generated. Fourteen students treated the expression $t + u$ as the unknown and directly solved the equation (SP1); 2 solved the equation after substituting x for $t + u$ (SP2); and 3 solved the equation by expanding the left side of the equation first (SP3). Two students produced the correct solution by trying the item options, which is a strategy unrelated to Attribute A6, “solving for a linear equation.” Hence, this strategy, which leads to the correct answer, is inconsistent with our attribute description.

Attribute A7 requires the “skills of setting up and solving for a quadratic equation, which generally involves the skills in solving a linear equation and additional skills (e.g., factoring).” Thus, Attribute A7 is considered a more complex attribute than A6. For example, in Item 16 (“For what positive number is the square root of the number the same as the number divided by 40?”), 12 out of the 21 students in our study correctly answered the item (see Figure 13 on next page). Of these 12 students, 9 adopted the strategy of solving the quadratic equation. Depending on how the students solved the equation, three different solution paths were identified. One student solved the equation by multiplying x^{-1} on both sides of the equation (SP1); 7 students, by squaring both sides of the equation (SP2); and 1, by generating the answer directly from the equation (SP3). The remaining 3 students adopted the strategy of plugging in numbers. Again, it is important to note that although 3 students produced the correct answer by plugging in numbers, their strategy was unrelated to our inferences about their mastery of Attribute A7, “setting up and solving for a quadratic equation.” Hence, this strategy, which leads to the correct answer, is inconsistent with our attribute description.

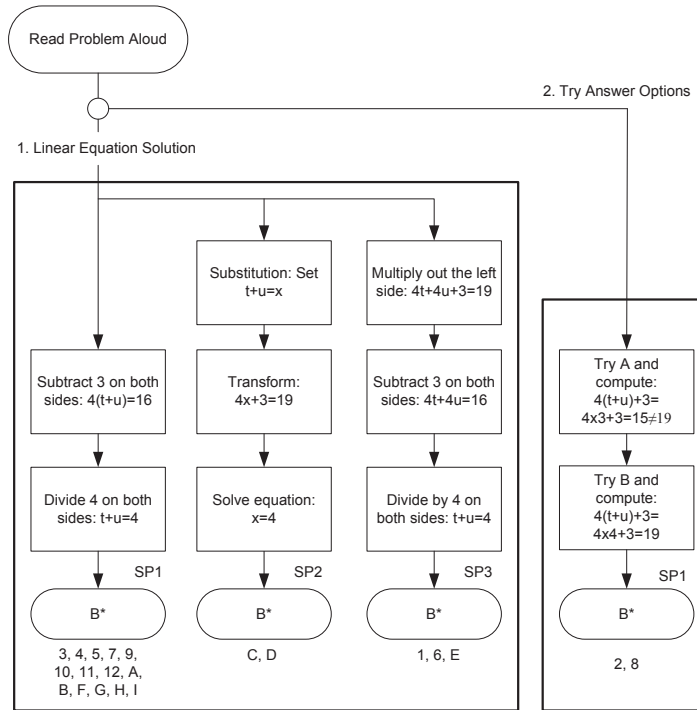


Figure 12. Problem-solving strategies for Attribute 6 in Hierarchy 2.

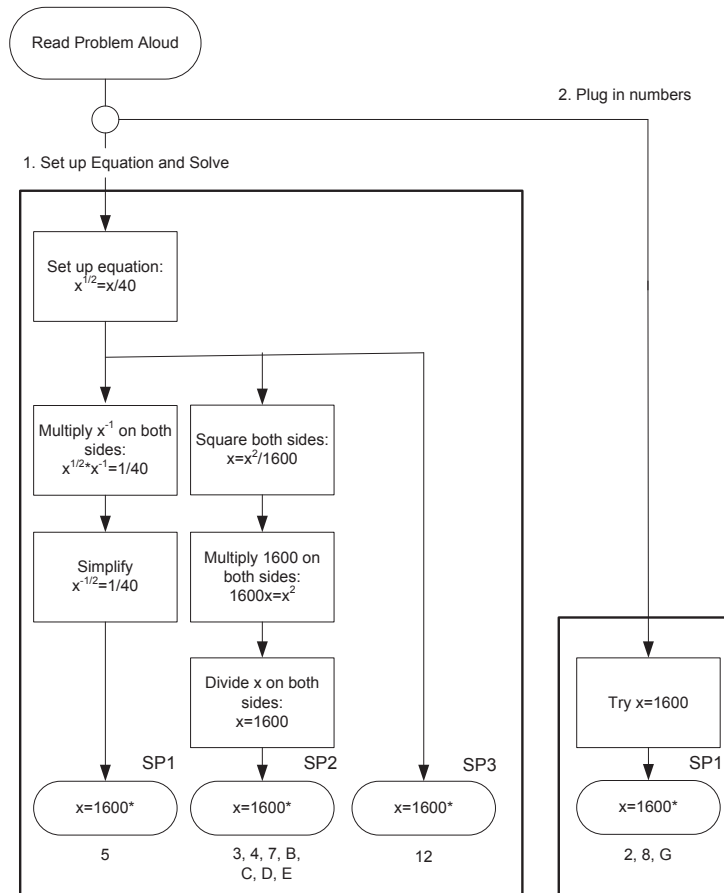


Figure 13. Problem-solving strategies for Attribute 7 in Hierarchy 2.

The fourth branch deals with the skills involved in functional graph reading. Attribute A8 represents “the skills of mapping a graph of a familiar function (e.g., a parabola) with its corresponding function.” This attribute involves the “knowledge about the graph of a familiar function and/or substituting points in the graph.” For example, in Item 4, the student needs to visually examine a graph or find random points in the graph and substitute the points in the equation of the function to find a match between the graph and the function. Sixteen students correctly answered this item. In solving the item, two strategies were used: visual inspection and substitution of random points (see Figure 14). Of the 16 students, 11 used the first strategy, while the remaining 5 students used the second strategy. Of the 11 students who used the first strategy, two solution paths were generated. Seven students produced the answer by observing the graph and eliminating the wrong options (SP1), and 4 students produced the answer by finding the relationship between the graph and the graph of $y = x^2$ (SP2).

Attribute A9, on the other hand, deals with the “abstract properties of functions, such as recognizing the graphical representation of the relationship between independent and dependent variables.” The graphs of less familiar functions, such as a periodic function or a function of higher-power polynomials, may be involved. Therefore, Attribute A9 is considered more difficult than Attribute A8. For example, in Item 14, which measures Attribute A9, the graph for a higher-power polynomial is used. To solve this item, the student needs to recognize the equivalent relationship between $f(x)$ and y , and that the number of times the graph crosses the line $y = 2$ produces the number of values of x that make $f(x) = 2$. Fifteen students correctly answered this item. In solving the item, two strategies were used: drawing lines across the graph and visual inspection (see Figure 15 on the next page). Of the 15 students, 7 used the first strategy, while the remaining 8 students used the second strategy.

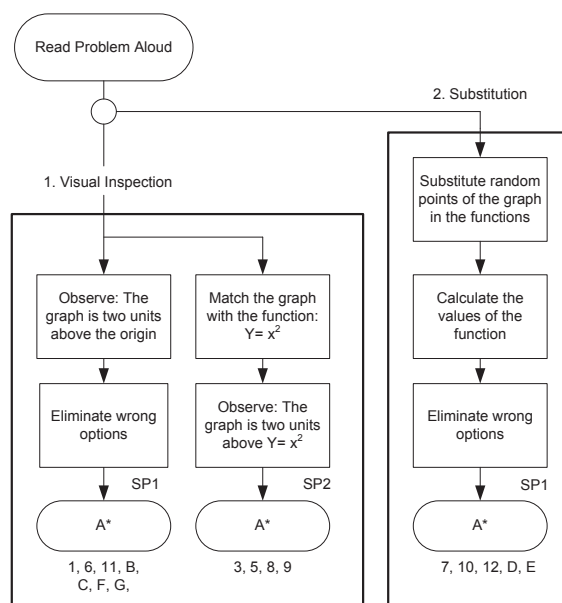


Figure 14. Problem-solving strategies for Attribute 8 in Hierarchy 2.

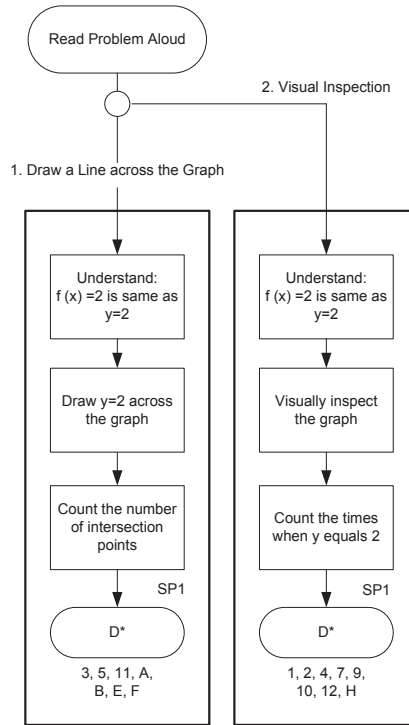


Figure 15. Problem-solving strategies for Attribute 9 in Hierarchy 2.

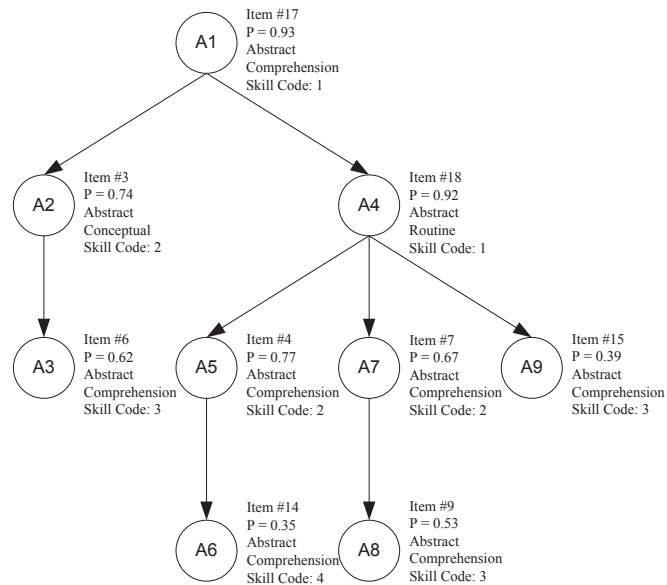


Figure 16. Ratios and algebra, Hierarchy 3.

Hierarchy 3 presents a cognitive model of task performance for skills in the general area labeled ratios and algebra. The hierarchy contains two independent branches that share a common prerequisite, Attribute A1. The first branch includes two additional attributes, A2 and A3, and the second branch includes a self-contained subhierarchy that includes Attributes A4 through A9.

Three independent branches compose the subhierarchy: Attributes A4, A5, and A6; Attributes A4, A7, and A8; and Attributes A4 and A9. The HCI_i for this model in Gierl, Wang et al. (2007) was 0.80, indicating strong model–data fit. The complete set of attributes used in Hierarchy 3 is summarized in Table 4.

Table 4

Summary of the Attributes Required to Solve the Items in Hierarchy 3, Ratios and Algebra

Attribute A1 represents the most basic arithmetic operation skills (e.g., addition, subtraction, multiplication, and division of numbers).
Attribute A2 includes knowledge about the properties of factors.
Attribute A3 involves the skills of applying the rules of factoring.
Attribute A4 includes the skills required for substituting values into algebraic expressions.
Attribute A5 represents the skills of mapping a graph of a familiar function (e.g., a parabola) with its corresponding function—see also Hierarchy 2, Attribute 8.
Attribute A6 deals with the abstract properties of functions, such as recognizing the graphical representation of the relationship between independent and dependent variables—see also Hierarchy 2, Attribute 9.
Attribute A7 requires the skills to substitute numbers into algebraic expressions.
Attribute A8 represents the skills of advanced substitution. Algebraic expressions, rather than numbers, need to be substituted into another algebraic expression.
Attribute A9 is related to skills associated with rule understanding and application.

As a prerequisite attribute, A1 represents “basic arithmetic skills with operations (e.g., addition, subtraction, multiplication, and division of numbers).” Item 17 serves as an example to illustrate this attribute. In this item, given $4(t + u) + 3 = 19$, one needs to subtract 3 from 19 and then divide 16 by 4 to solve for $(t + u)$. All

21 students correctly answered the item. Three strategies were involved: arithmetic operation, linear equation solution, and trial of options (see Figure 17). Seventeen of the 21 students adopted the first strategy, 2 adopted the second strategy, and the remaining 2 adopted the third strategy. Of the 17 students who used the first strategy,

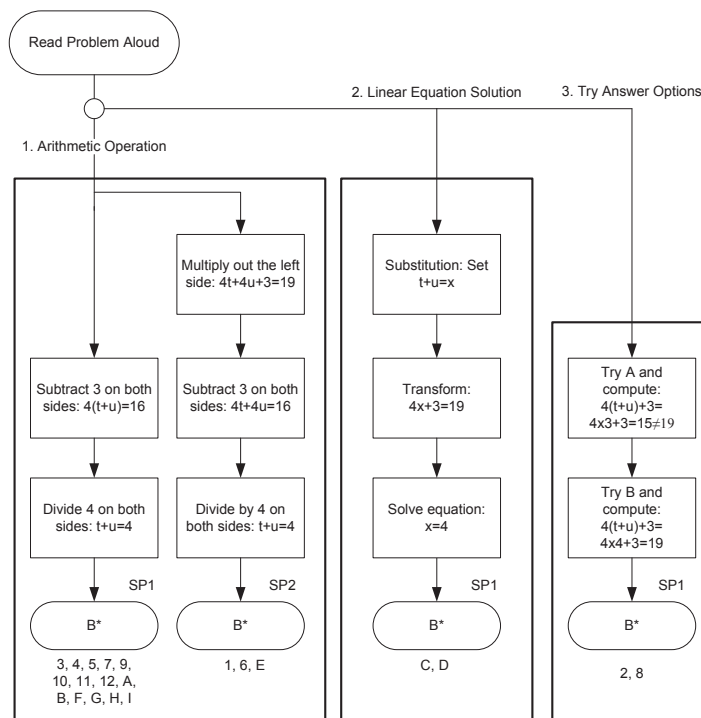


Figure 17. Problem-solving strategies for Attribute 1 in Hierarchy 3.

14 students used SP1 and the remaining 3 used SP2, depending on the order of arithmetic operations.

Attributes A2 and A3 both deal with factors. Attribute A2 includes “knowledge about the properties of factors.” For example, in Item 3, which measures Attribute A2, the student must recognize the property that the value of at least one factor must be zero if the product of multiple factors is zero. Sixteen of the 21 students correctly answered this item. Two strategies were adopted: applying knowledge of factors and plugging in numbers (see Figure 18). Thirteen of the 16 students used the first strategy, while the remaining 3 students used the second strategy. However, the second strategy—plugging in numbers—does not reflect the skills associated with the attribute of “knowledge about the properties of factors.”

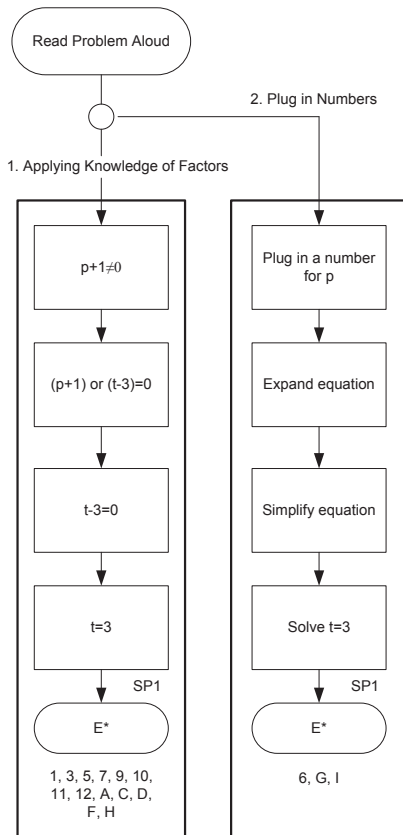


Figure 18. Problem-solving strategies for Attribute 2 in Hierarchy 3.

In addition to the knowledge about properties of factors in Attribute A2, A3 involves “the skills of applying the rules of factoring.” Item 6 measures Attribute A3. For this item, students are required to factor $\frac{9x + 9y}{10a - 10b}$ into the product of $\frac{x + y}{a - b}$ and $\frac{9}{10}$ to obtain a correct answer.

Nineteen students correctly answered Item 6. Three strategies were used: applying the rules of factoring, plugging in random numbers, and solving the equation (see Figure 19). Of the 19 students who correctly answered this item, 14 adopted the first strategy, 4 adopted the second strategy, and the remaining 1 adopted the third strategy. However, the second strategy, plugging in random numbers, does not reflect the skills associated with those that are measured by this attribute.

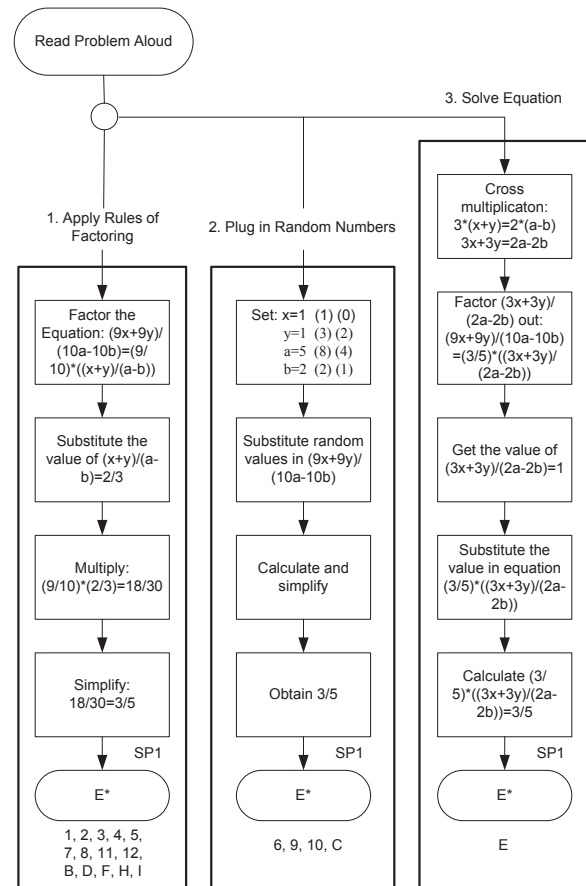


Figure 19. Problem-solving strategies for Attribute 3 in Hierarchy 3.

The self-contained subhierarchy contains six attributes. Among these attributes, Attribute A4 is the prerequisite for all other attributes in the subhierarchy. Attribute A4 has A1 as a prerequisite because A4 not only represents basic skills in arithmetic operations (i.e., Attribute A1), but it also involves the substitution of values into algebraic expressions, which is more abstract and therefore more difficult than Attribute A1. For instance, in Item 18, the examinee needs to substitute an algebraic expression or the values of variables to compute the value of another variable. All 21 students correctly answered the item. One dominant strategy, substitution, was adopted by all students (see Figure 20). Depending on the order of substitution, two solution paths were identified for the dominant strategy of substitution. Twenty students substituted the values of variables consecutively into the algebraic expressions to obtain the final answer (SP1). The remaining student substituted an algebraic expression into another algebraic expression first and then substituted the values of variables to obtain the final answer (SP2).

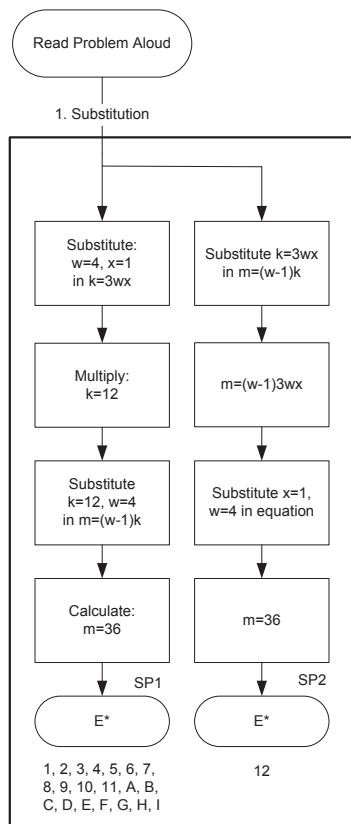


Figure 20. Problem-solving strategies for Attribute 4 in Hierarchy 3.

The first branch in the subhierarchy, which contains Attributes A5 and A6, deals mainly with functional graph reading. Attribute A5 representing the “skills of mapping a graph of a familiar function (e.g., a parabola) with its corresponding function,” and Attribute A6, which deals with the “abstract properties of functions,” were illustrated in Hierarchy 2.

The second branch in the subhierarchy considers the skills associated with advanced substitution. Attribute A7 requires “the skills to substitute numbers into algebraic expressions.” The complexity of Attribute A7 relative to Attribute A4 lies in the concurrent management of multiple pairs of numbers and multiple equations. For example, in Item 7, examinees are asked to identify which function matches the pairs of x and y values. To solve this item, the examinee needs to substitute three pairs of x and y values into the five functions provided to find the correct pair. Twenty out of 21 students correctly answered the item. Two strategies were adopted: multiple substitution and pattern recognition (see Figure 21). Nineteen out of the 20 students adopted the first strategy and obtained the correct answer by substituting the number pairs in the functions provided in the answer options. The remaining student obtained the correct answer by recognizing the pattern implied by the number pairs and then matching the pattern with the functions provided in the answer options.

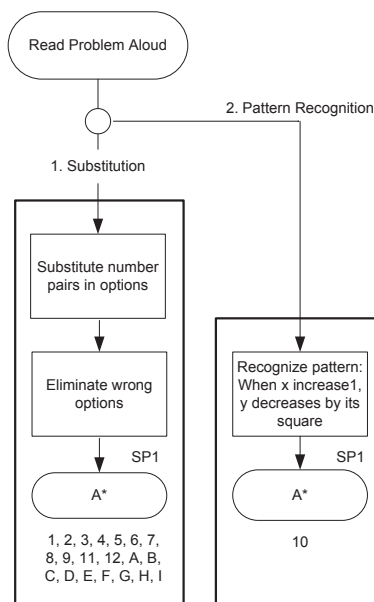


Figure 21. Problem-solving strategies for Attribute 7 in Hierarchy 3.

Attribute A8 also represents “the skills of advanced substitution.” However, what makes Attribute A8 more difficult than Attribute A7 is that “algebraic expressions, rather than numbers, need to be substituted into another algebraic expression.” For instance, in Item 9, the examinee is given $x = 3v$, $v = 4t$, and $x = pt$, and then asked to find the value of p . Examinees need to substitute x and v into the equation, set up an equation such as $x = 3v = 12t = pt$, and then substitute a numeric value for t (such as 1) and for v (such as 4) to produce $p = 12$. Nineteen out of 21 students correctly answered the item. Two strategies were adopted: substitution and plugging in numbers (see Figure 22). Fourteen students adopted the first strategy, and 5 students adopted the second strategy. The 5 students who produced the correct answer by plugging in random numbers used a strategy unrelated

to our inferences about their mastery of Attribute A8, skills of substitution. Hence, this strategy, which leads to the correct answer, is inconsistent with our attribute description.

The last branch in the subhierarchy contains only one additional attribute, A9, related to “skills associated with rule understanding and application.” In Item 15, for example, examinees are presented with $x \Delta y = x^2 + xy + y^2$, and then asked to find the value of $(3 \Delta 1) \Delta 1$. To solve this item, the examinee must first understand what rule Δ represents, and then twice substitute the rule into the expression $(3 \Delta 1) \Delta 1$ to produce the solution. Eighteen out of 21 students correctly answered the item, and they adopted one dominant strategy: understanding and application of the rule (see Figure 23).

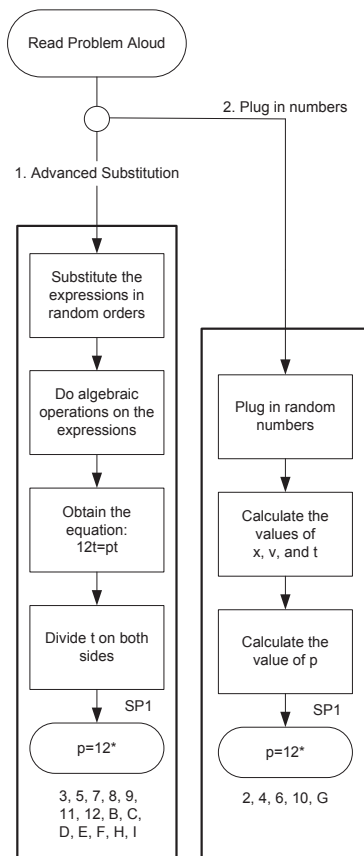


Figure 22. Problem-solving strategies for Attribute 8 in Hierarchy 3.

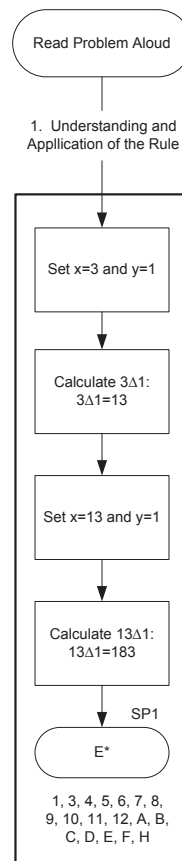


Figure 23. Problem-solving strategies for Attribute 9 in Hierarchy 3.

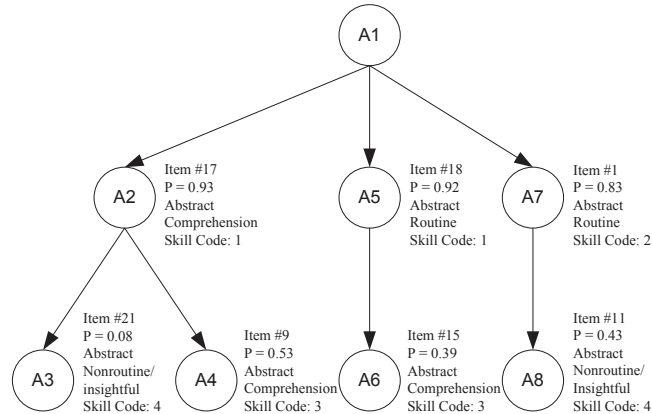


Figure 24. Equation and inequality solutions, algebraic operations, algebraic substitution, and exponents, Hierarchy 4.

Hierarchy 4 serves as a cognitive model of task performance for skills in a diverse number of areas, including equation and inequality solutions, algebraic operations, algebraic substitution, and exponents. The hierarchy contains three independent branches, which share a common prerequisite, Attribute A1. The first branch includes a subhierarchy composed of three attributes: A2, A3, and A4. There are two branches in this subhierarchy: Attributes A2 and A3 and Attributes

A2 and A4. Aside from Attribute A1, both the second and the third branch in Hierarchy 4 include two additional attributes—A5 and A6 for Branch 2 and A7 and A8 for Branch 3. The HCI_i for this model in Gierl, Wang et al. (2007) was 0.92, indicating excellent model–data fit—the highest HCI_i value, in fact, of the four algebra models in our study. The complete set of attributes used in Hierarchy 4 is summarized in Table 5.

Table 5

Summary of the Attributes Required to Solve the Items in Hierarchy 4, Equation and Inequality Solutions, Algebraic Operations, Algebraic Substitution, and Exponents

Attribute A1 includes basic language knowledge enabling students to understand the test item and basic mathematical knowledge and skills, such as the property of absolute value and arithmetic operations—see also Hierarchy 2, Attribute 1.
Attribute A2 represents the most basic arithmetic operation skills (e.g., addition, subtraction, multiplication, and division of numbers)—see also Hierarchy 3, Attribute 1.
Attribute A3 involves the skills required for solving a quadratic inequality with two variables.
Attribute A4 represents the skills of solving multiple linear equations.
Attribute A5 considers the skills of substituting values into algebraic expressions—see also Hierarchy 3, Attribute A7.
Attribute A6 involves the skills of rule understanding and substitution—see also Hierarchy 3, Attribute A9.
Attribute A7 requires the basic knowledge of exponential and power addition operation—see also Hierarchy 2, Attribute A2.
Attribute A8 represents the knowledge of power multiplication and flexible application of multiple rules in exponential operation—see also Hierarchy 2, Attribute A3.

As a prerequisite, Attribute A1 includes “basic language knowledge enabling students to understand the test item, and basic mathematical knowledge and skills, such as the property of absolute value and arithmetic operations.” There are three attributes in the first branch of Hierarchy 4, with Attribute A2 serving as the prerequisite. Attribute A2 has A1 as a prerequisite because A2 not only represents basic skills in mathematic operations, it also involves the skills in producing a linear equation solution, which requires the management of basic mathematic skills. Attribute A2 was illustrated in Hierarchy 3.

The subhierarchy in the first branch contains one additional attribute: A3, which includes “the skills required for solving quadratic inequalities with two variables.” The prerequisite to A3 is A2, as A3 also includes the skills of expanding the square of sums of two variables and simplifying algebraic expressions. For example, in Item 21, one needs to expand $(x + y)^2 - (x - y)^2 \geq 25$ as $(x^2 + y^2 + 2xy) - (x^2 + y^2 - 2xy) \geq 25$ and then simplify it into $y^2 \geq \frac{25}{4}$. Moreover, examinees should recognize

that given $0 \leq x \leq y$, the least possible value of y equals x . Therefore, the inequality changes to $xy \geq \frac{25}{4}$. To correctly answer Item 21, students also need to take the square root on both sides of the inequality. Only two students correctly answered the item. Although both students used the same algebraic operation strategy, two different solution paths were used (see Figure 25). One student obtained the correct answer by spreading out the inequality, simplifying it, and using insight (SP1). The other student obtained the correct answer by simplifying the inequality (SP2). However, this student made an important mistake in simplifying the inequality, which led, ironically, to the correct answer. Therefore, this student's answer cannot be used to infer that he or she possesses the cognitive skills required to master this attribute, even though he or she identified the correct answer.

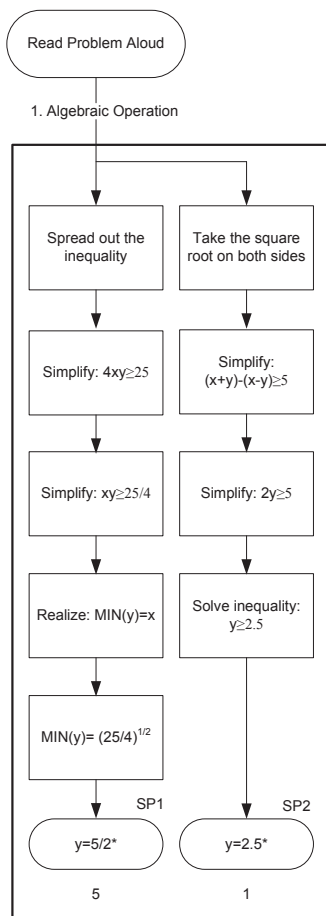


Figure 25. Problem-solving strategies for Attribute 3 in Hierarchy 4.

The second branch in the subhierarchy consists of one additional attribute, A4, which represents “the skills of solving multiple linear equations.” What makes A4 more difficult than A2 is the concurrent management of multiple linear equations, rather than a single linear equation. Moreover, the solution involves operations on algebraic expressions rather than simply numeric values and is therefore more abstract and difficult. For example, in Item 9, given $x = 3v$, $v = 4t$, $x = pt$, the examinees need to find the value of p . One way to approach the item is to solve for p first using x and t such that $p = \frac{x}{t}$. Then the examinee needs to solve for t given v such that $t = \frac{v}{4}$.

The values of x and t , both in terms of v , can be substituted into $p = \frac{x}{t}$ to get the answer, $p = 12$. Nineteen out of 21 students correctly answered the item. Two strategies were adopted: equation operation and plugging in numbers (see Figure 26). Fourteen students adopted the first strategy, and 5 students adopted the second strategy. As in the previous examples, plugging in random numbers is not a strategy that is consistent with the attribute probed by this item. Hence, this strategy, which leads to the correct answer, is inconsistent with our attribute description.

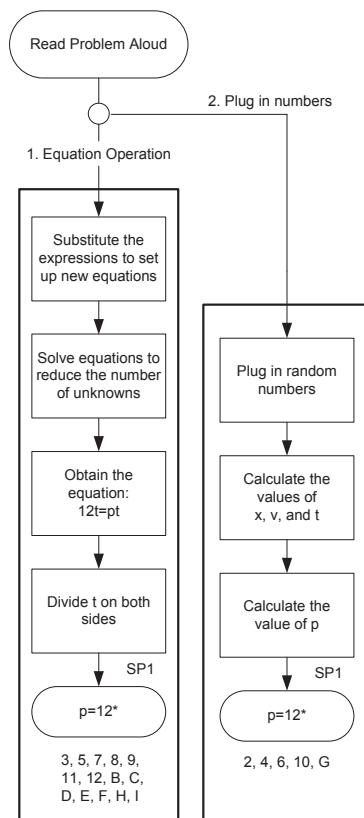


Figure 26. Problem-solving strategies for Attribute 4 in Hierarchy 4.

The second branch in Hierarchy 4 deals with the skills on substitution. Attribute A5 considers “the skills of substituting values into algebraic expressions.” A6 involves “the skills of rule understanding and substitution.” These attributes were illustrated in Hierarchy 3.

The third branch considers basic exponential operations. Attribute A7 represents “the basic knowledge of exponential and power addition,” and Attribute A8 includes skills related to “multiplication operations and flexible application of multiple rules in exponential operation.” Attributes A7 and A8 were illustrated in Hierarchy 2.

Summary and Implications for Diagnostic Assessment Using the AHM

Summary of Current Study

The AHM for cognitive assessment is a psychometric method for classifying examinees’ test item responses into a set of structured attribute patterns associated with different components that are specified in a cognitive model of task performance. Cognitive diagnostic assessment must be informed by empirical studies on domain knowledge and skill acquisition so that psychological theory can be linked to measurement practice, thereby promoting cognitive inferences about examinees’ performance. The AHM attempts to forge this link by using a cognitive model to guide both the development of items and the interpretation of test scores. As diagnostic assessments continue to develop, they must also be informed by innovative new psychometric procedures that can measure performance and improve methods for reporting this performance in ways that are consistent with contemporary psychological theory.

The purpose of the current study was to validate the cognitive models presented in Gierl, Wang et al. (2007), which initially were used to evaluate the cognitive problem-solving skills used by students to solve a subset of algebra items taken from the March 2005 administration of the SAT using the AHM. The cognitive models used in Gierl, Wang et al. were developed by content specialists but never validated using a sample of students from the target population. To address this problem, student response data were collected with verbal think-aloud methods to evaluate the knowledge structures and processing skills used by a sample of SAT test-takers to solve the algebra items. The verbal protocol data were collected in November 2005 by asking 21 students to think aloud as they solved the sample of 21 algebra items used in Gierl, Wang et al. We began this report by defining the phrase “cognitive model in educational measurement,” and by explaining why these models are important in the development and analysis of diagnostic assessments. Then we described the methods and analyses used to collect the verbal protocol data. Finally, we presented the results from our protocol analysis. In the final section, we highlight the implications of our results for making cognitive diagnostic inferences using the AHM.

Implications of Verbal Protocol Results for Diagnostic Assessment with the AHM

The verbal reports results have implications for model modifications and for producing diagnostic inferences with assessment methods like the AHM.

Model Modifications

The cognitive model of task performance used to develop the four algebra attribute hierarchies for the analyses presented in Gierl, Wang et al. (2007) were generated from a task analysis of the SAT Algebra I and II items to identify the mathematical concepts, operations, procedures, and strategies that students might use to solve items on the SAT. However, the evidence in Gierl, Wang et al. for supporting the algebra models was limited because it was only based on a task analysis of the items. To address this limitation, the four algebra attribute hierarchies were validated using student response data that had been collected with verbal think-aloud methods.

For the most part, the content-based cognitive models in algebra provided a good approximation to the student results. The HCI_i fit indices were moderate

to strong, ranging from 0.76 for Hierarchy 1 (revised) to 0.92 for Hierarchy 4. Only one model modification—a structural change—was made in one of the four cognitive models based on our analyses of the student response data. Initially, Attributes A4 and A5 were viewed as hierarchically independent. However, the results from the protocol analysis revealed that the skills involved in A5 require a higher level of cognitive skill than those involved in A4: For A5, students had to conduct an analysis on the conditions given in the item, while for A4 students only needed to apply basic algebraic skills. Therefore, the results from the protocol analysis led to a modification of Hierarchy 1, where a dependency between Attributes A4 and A5 was added.

Attribute consistency in the context of algebra problem solving is another issue raised by the results in our study. For example, in Hierarchy 1, Attribute A4 is labeled “representing and executing multiple basic algebraic skills.” The multiple skills required in this attribute apply to ratio problems because the items in this hierarchy contain ratios. Hence, this skill requires a number of simultaneous steps, including the ability to apply such operations as square, square root, multiplication, and division. Twelve out of the 21 students answered Item 16 (see Figure 6). Of these 12 students, 9 adopted the strategy

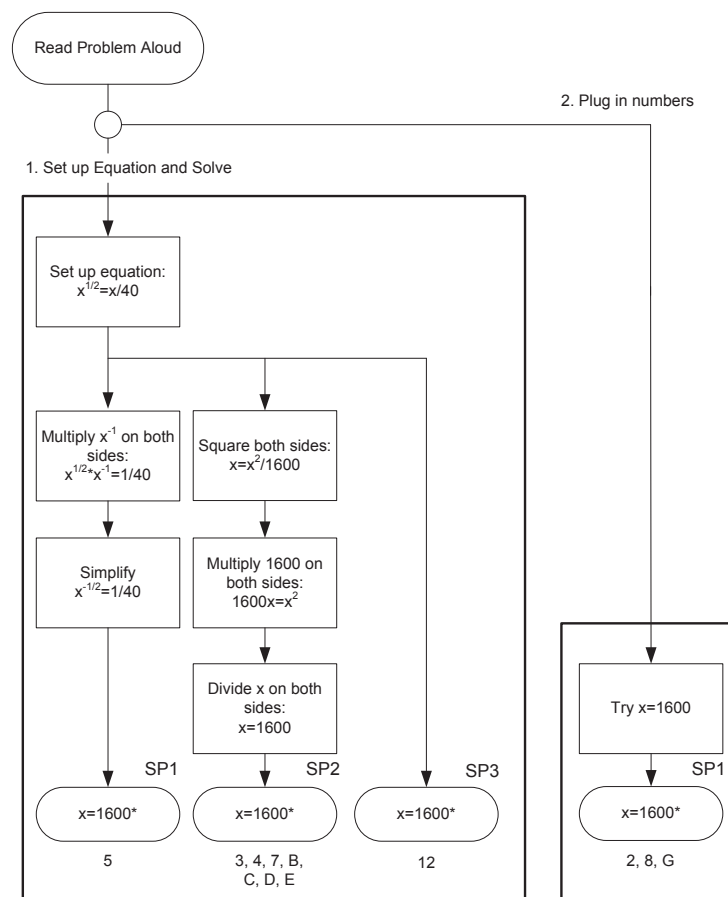


Figure 6. Problem-solving strategies for Attribute 4 in Hierarchy 1.

of solving a quadratic equation. Depending on how the students solved the equation, three different solution paths were identified. One student solved the equation by multiplying x^{-1} on both sides of the equation (SP1); 7 students, by squaring both sides of the equation (SP2); and 1 by generating the answer directly from the equation (SP3). The remaining 3 students adopted the strategy of plugging in numbers to the constructed-response item.

Hierarchy 2, Attribute A7 also contains Item 16. But in the context of this set of items, A7 is deemed to measure the skills of setting up and solving for a quadratic equation, which generally involve the skills in solving a linear equation. Hence, the cognitive skills in Attribute A7, Hierarchy 2 can be described more specifically than the cognitive skills measured by Attribute A4, Hierarchy 1, in part, because Hierarchy 2 (nine attributes in total) is more complex than Hierarchy 1 (five attributes in total) but also because A7 (in Hierarchy 2) has more prerequisite attributes than A4 (in Hierarchy 1)—two prerequisite attributes versus one prerequisite attribute, respectively.

This finding reveals that attribute grain size is constantly an issue of concern in diagnostic testing, as it must be monitored and consistently applied when labeling and interpreting attributes. Gierl, Wang et al. (2007) noted that the hierarchy of attributes required to perform well in a domain should be identified *prior to developing the test items* with the AHM. Yet, in the current study no new items were developed for the cognitive models of task performance used to produce the algebra attribute hierarchies. Also, *item-* rather than *attribute-based* hierarchies were used (the distinction between item- and attribute-based hierarchies is described in more detail in the next section). While item-based hierarchies are convenient because test items and examinee response data are available, they are also limited because the cognitive model must be generated post hoc, and only existing items can be used to operationalize the attributes. One important consequence of this post-hoc approach is that the fit between the cognitive model and the item-based attributes is tenuous. As is apparent from the results in our study, the fit between the attributes and items was established, in some cases, by modifying our interpretation of the attribute measured by the item. One outcome of this post-hoc approach is that the same attributes were not always described in a consistent manner between different models; rather, the interpretation of the attribute was adjusted to suit the specific model in question. It must be noted, however, that this risk is inherent to any cognitive analysis of an existing test using retrofitting procedures because the items control the characteristics of the models; only with principled test design will the

model control the characteristics of the items, including the level of cognitive analysis. Therefore, to overcome this limitation associated with a cognitive post-hoc approach, principled test design is required. Using this strategy, the cognitive model of task performance is first identified and evaluated, and then the test items are developed to measure the attributes in the model.

Diagnostic Inferential Errors

The attribute hierarchy serves as a representation of the underlying cognitive model of task performance. These models provide the interpretative framework for guiding item development so that test performance can be linked to specific cognitive inferences about examinees' knowledge, processes, and strategies. These models also provide the means for connecting cognitive principles with measurement practices, in the spirit prescribed by Pellegrino, Baxter, and Glaser (1999):

...[I]t is the pattern of performance over a set of items or tasks explicitly constructed to discriminate between alternative profiles of knowledge that should be the focus of assessment. The latter can be used to determine the level of a given student's understanding and competence within a subject-matter domain. Such information is interpretative and diagnostic, highly informative, and potentially prescriptive. (p.335)

Hence, AHM analyses are predicated on the assumption that the attribute hierarchy is true.

To develop these models, we must also assume that student performance is goal directed, purposeful, and principled based on the instructional events that precede testing. Students are not expected to guess, plug in numbers from the multiple-choice alternatives to incomplete equations and expressions, or randomly apply option alternatives to information in the multiple-choice stem. We must make these assumptions because random performance is impossible to predict and, therefore, model. Moreover, random performances, even when they do lead to the correct answer, cannot inform instruction.

Unfortunately, as the results of our study make clear, our assumption about purposeful student performance is not always accurate. This is because students are motivated to produce the right answer even by the wrong means, and because the multiple-choice item format permits guessing. Four strategies not taken into account by our cognitive models were used to correctly solve algebra items: plugging in numbers, guessing, using the calculator, and trying answer options. A summary of the prevalence of these strategies for each hierarchy is presented in Table 6.

Table 6

Summary of the Strategies Used to Correctly Solve Items but Excluded from the Cognitive Models

<i>Hierarchy 1</i>		
Attribute (Item)	Strategy	Number of Students
A4 (16)	1. Plug in numbers	3 (1 Male; 2 Females)
A5 (20)	1. Plug in numbers	2 (2 Males)
	2. Guess	2 (2 Males)
<i>Hierarchy 2</i>		
Attribute (Item)	Strategy	Number of Students
A2 (1)	1. Use the calculator	1 (1 Female)
A3 (11)	1. Plug in numbers	6 (3 Males; 3 Females)
A6 (17)	1. Try answer options	2 (2 Males)
A7 (15)	1. Plug in numbers	3 (2 Males; 1 Female)
<i>Hierarchy 3</i>		
Attribute (Item)	Strategy	Number of Students
A1 (17)	1. Try answer options	2 (2 Males)
A2 (3)	1. Plug in numbers	2 (1 Male; 1 Female)
A3 (6)	1. Plug in numbers	4 (3 Males; 1 Female)
A8 (9)	1. Plug in numbers	3 (2 Males; 1 Female)
<i>Hierarchy 4</i>		
Attribute (Item)	Strategy	Number of Students
A2 (17)	1. Try answer options	2 (2 Males)
A4 (9)	1. Plug in numbers	3 (2 Males; 1 Female)
A8 (11)	1. Plug in numbers	6 (3 Males; 3 Females)

Although the number of strategies excluded from our cognitive models is not large and the strategies' use is infrequent, these problem-solving approaches will produce errors in our diagnostic inferences because we must assume that students possess the attributes outlined in the cognitive model if they produce a correct response. That is, we assume that the correspondence between the cognitive model and the response outcome is perfect. One purpose of the current study was to evaluate this assumption using SAT items and examinees. Our results revealed that the algebra models of task performance provide an acceptable approximation to the cognitive skills initially identified by content specialists and used by students to solve the 21 algebra items. But we also acknowledge that the correspondence between the cognitive model and the response outcome is not perfect. How can this erroneous assumption be addressed?

The first solution requires that we begin by defining the cognitive model of task performance and then generate items systematically using the reduced incidence matrix from the AHM analysis to measure each attribute. Because we retrofit existing test items to the cognitive model, each cognitive model in our analysis can be

described as an *item-based* algebra hierarchy. Gierl, Wang et al. (2007) claimed:

This type of hierarchy uses the test item as the unit of analysis. An item-based hierarchy can be compared to an attribute-based hierarchy where the attribute is the unit of analysis. Item-based hierarchies are typically generated when cognitive models are "retrofit" to existing data. While these types of hierarchies are convenient because examinee response data are available, they can also be very limiting if an appropriate cognitive model cannot be identified to describe examinee performance and/or if a small number of items are used in the model thereby decreasing the reliability of each attribute measured. (p. 12)

Our retrofitting approach clearly limited the number of items we could use to measure each attribute: We identified one item per attribute in the current study. To overcome this limitation, principled test design could be used to specify the cognitive model and then create multiple, replicable test items to systematically measure each attribute in the model.

If the item-based hierarchy is to be maintained, a second solution is to simply increase the number of items measuring each attribute. This approach may yield less inferential error because a larger sample of examinee behavior would be available for each attribute (i.e., three items per attribute provide a broader sample of the examinees' cognitive skills than one item per attribute). The intention in sampling the same cognitive skills on multiple test items is that the anomalous strategies we encountered—plugging in numbers, guessing, using a calculator, and trying answer options—would likely not lead to the correct solution consistently. As a result, the statistical pattern recognition approach we used to produce the attribute probabilities would yield a lower value for examinees who use these anomalous strategies. Unfortunately, this approach would also decrease the value of HCI_i -fit indices for the model because more compromises—and, hence, poorer-fitting items—would need to be associated with each attribute, given that the distribution of items is likely to be uneven across attributes in the hierarchy. These compromises are necessary because algebra items on the SAT were not developed with an explicit cognitive model of task performance. This important compromise suggests that our first solution—principled test design—is preferred.

The Evolution of Cognitive Models

A cognitive model in educational measurement refers to a “simplified description of human problem solving on standardized educational tasks, which helps to characterize the knowledge and skills students at different levels of learning have acquired and to facilitate the explanation and prediction of students' performance” (Leighton and Gierl, 2007a, p. 5). These models provide an interpretative framework to guide item development so that test performance can be linked to specific cognitive inferences about examinees' knowledge, processes, and strategies. Recently, Mislevy (2006) described six aspects or steps in model-based reasoning in science. These six steps, summarized in Table 7, provide an excellent framework for considering our progress in developing cognitive models in algebra on the SAT.

The first step is model formation. The researcher must establish a correspondence between some real-world phenomenon and a model. The empirical considerations for modeling cognitive skills using the AHM with hierarchical structures are described in Leighton et al. (2004) and Gierl, Wang et al. (2007). The psychological considerations for modeling cognitive skills using psychometric methods and linking these skills to diagnostic inferences are outlined in Leighton and Gierl (2007a). The second aspect is model elaboration. In this step, models are developed and detailed. Over the course of two studies—Gierl, Wang et al. (2007) and the present study—we have developed four cognitive models

of algebra performance that describe different aspects of problem solving using subsets of items from Algebra I and II. These four models were elaborated using results from task analyses conducted by content specialists, and from verbal think-aloud protocols by SAT examinees using Algebra I and II items. Although the models have similarities (i.e., some models share attributes and items) and differences, they provide a concise yet detailed description of the types of skills that could be evaluated on the SAT. The third aspect, model use, provides structure to the model so that explanations and predictions can be made. By ordering the algebra attributes within a hierarchy of cognitive skills, our model specifies how the attributes are structured internally by SAT examinees when they solve test items.

Model evaluation is the fourth step. Here, the correspondence between the model components and their real-world counterparts is assessed. The purpose of the current study was to evaluate four plausible cognitive models of algebra performance by comparing representations of content specialists and SAT examinees in order to establish the correspondence between the model and examinees' problem-solving procedures. In step five, model revisions can occur. Our evaluation of the cognitive models in Figure 1 using student response data from verbal reports led to one key structural change in Hierarchy 1. But we also noted that the content-based cognitive models in algebra provided an excellent approximation to the student results. Finally, in step six, *model-based inquiry* can occur. In this step, the model is applied to student response data, where outcomes and actions are guided by model-based inferences. In other words, when steps one through five have been satisfied, the model can be used in step six. The types of model-based inferences that can be produced by the AHM in algebra were illustrated in Gierl, Wang et al. (2007) using Hierarchy 3 with a random sample of students who took the March 2005 administration of the SAT.

Taken together, results from both the current study and from past studies (e.g., Cui, Leighton, Gierl, and Hunka, 2006; Gierl, in press; Gierl, Zheng, and Cui, in press; Gierl, Cui, and Hunka, 2006; Gierl, Leighton, and Hunka, 2000; Gierl, Leighton, and Hunka, 2007; Leighton and Gierl, 2007a; Gierl, Tan, and Wang, 2005; Leighton, Gierl, and Hunka, 2004; VanderVeen, Huff, Gierl, McNamara, Louwerson, and Graesser, 2007; Wang and Gierl, 2006; Wang, Gierl, and Leighton, 2006) reveal that an empirically based body of evidence now exists to justify and support the use of the AHM for making diagnostic inferences on the mathematics section of the SAT.

Table 7

Six Steps in Model-Based Reasoning

Model Formation	Establishing a correspondence between some real-world phenomenon and a model, or abstracted structure, in terms of entities, relationships, processes, etc. Includes establishing the scope and grain size to the model and determining which aspects of the phenomenon to include and exclude.
Model Elaboration	Combining, extending, and adding detail to a model, and establishing correspondences across overlapping models. Often done by assembling smaller models into larger assemblages or fleshing out more general models with details.
Model Use	Reasoning through the structure of a model to make explanations, predictions, conjectures, etc.
Model Evaluation	Assessing the correspondence between the model components and their real-world counterparts, with emphasis on anomalies and important features not accounted for by the model.
Model Revisions	Modifying and elaborating a model for a phenomenon in order to establish a better correspondence. Often initiated by model evaluation procedures.
Model-Based Inquiry	Working interactively between phenomena and models, using all of the previous steps. Emphasis on monitoring and taking actions with regard to model-based inferences vis-à-vis real-world feedback.

References

- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355–365.
- Cui, Y., Leighton, J. P., Gierl, M. J., & Hunka, S. (2006, April). *A person-fit statistic for the attribute hierarchy method: The hierarchical consistency index*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Dawson, M. R. W. (1998). *Understanding cognitive science*. Malden, MA: Blackwell.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: The MIT Press.
- Gierl, M. J. (2007). Making diagnostics inferences about cognitive attributes using the rule space model and attribute hierarchy method. *Journal of Educational Measurement*, *44*, 325–340. Invited paper appearing in special issue on IRT-based cognitive diagnostic models and related methods, Lou DiBello & William Stout (guest editors).
- Gierl, M. J., Leighton, J. P., & Hunka, S. (2000). Exploring the logic of Tatsuoka's rule-space model for test development and analysis. *Educational Measurement: Issues and Practice*, *19*, 34–44.
- Gierl, M. J., Leighton, J. P., & Hunka, S. (2007). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 242–274). Cambridge, UK: Cambridge University Press.
- Gierl, M. J., Cui, Y., & Hunka, S. (2006). The attribute hierarchy method for cognitive assessment: Technical developments. Manuscript submitted for publication.
- Gierl, M. J., Wang, C., & Zhou, J. (2007, July). *Using the attribute hierarchy method to develop cognitive models and evaluate problem-solving skills in algebra on the SAT*. New York: The College Board.
- Gierl, M. J., Tan, X., & Wang, C. (2005). *Identifying content and cognitive dimensions on the SAT* (Research Report No. 2005-11). New York: The College Board.
- Gierl, M. J., Zheng, Y., & Cui, Y. (2008). Using the attribute hierarchy method to identify and interpret the cognitive skills that produce group differences. *Journal of Educational Measurement*, *45*, 65–89.
- Kuhn, D. (2001). Why development does (and does not) occur: Evidence from the domain of inductive reasoning. In J. L. McClelland & R. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 221–249). Hillsdale, NJ: Erlbaum.
- Leighton, J. P. (2004). Avoiding misconceptions, misuse, and missed opportunities: The collection of verbal reports in educational achievement testing. *Educational Measurement: Issues and Practice*, *23*, 6–15.
- Leighton, J. P., & Gierl, M. J. (2007a). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, *26*, 3–16.
- Leighton, J. P., & Gierl, M. J. (2007b). Verbal reports as data for cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 146–172). Cambridge, UK: Cambridge University Press.
- Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, *41*, 205–236.
- Luecht, R. M. (2007). Using information from multiple-choice distractors to enhance cognitive-diagnostic score reporting. In J. Leighton & M. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education: Theory and Application* (pp. 319–340). Cambridge, UK: Cambridge University Press.
- Mislevy, R. J. (2006). Cognitive psychology and educational measurement. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 257–305). Westport, CT: Praeger.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3–62.
- Norris, S. P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. *Journal of Educational Measurement*, *27*, 41–58.
- O'Callaghan, R., Morley, M., & Schwartz, A. (2004, April). *Developing skill categories for the SAT math section*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.
- Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the “two disciplines” problems: Linking theories of cognition and learning with assessment and instructional practices. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education* (pp. 307–353). Washington, DC: American Educational Research Association.
- Reddy, R. N., & Ziegler, C. A. (1994). *Fortran 77 with 90: Applications for scientists and engineers* (2nd ed.). Minneapolis: West.
- Royer, J. M., Cisero, C. A., & Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research*, *63*, 201–243.
- Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York: Macmillan.
- Taylor, K. L., & Dionne, J. P. (2000). Accessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, *92*, 413–425.

-
- VanderVeen, A., Huff, K., Gierl, M., McNamara, D. S., Louwerse, M., & Graesser, A. (2007). Developing and validating instructionally relevant reading competency profiles measured by the critical reading section of the SAT Reasoning Test™. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 137–172). New York: Taylor & Francis.
- Wang, C., & Gierl, M. J. (2006, August). *Investigating the reading skill hierarchies on an L2 English test*. Paper presented at the Third International Conference on Speech, Writing, and Context Language and Literacy Acquisition, Faculty of Education, University of Alberta, Edmonton, Alberta, Canada.
- Wang, C., Gierl, M. J., & Leighton, J. P. (2006, April). *Investigating the cognitive processes underlying student performance on a foreign language reading test: An application of the attribute hierarchy method*. Poster presented at the annual meeting of the American Educational Research Association, San Francisco, CA.



CollegeBoard

inspiring minds™