Development of Cognitive Models in Mathematics to Promote Diagnostic Inferences about Student Performance

Mary Roduta Roberts
Cecilia Brito Alves
Andrea Gotzmann
Mark J. Gierl

Centre for Research in Applied Measurement and Evaluation

University of Alberta

Paper Presented at the Paper Discussion Session

"Psychometric Applications Including Instrument Development and Performance Prediction"

Annual meeting of the American Educational Research Association

San Diego, CA

April 13, 2009

Abstract

The purpose of this paper is to present one method for developing a cognitive model in Mathematics to promote diagnostic inferences about student performance. The cognitive model is the fundamental basis for developing diagnostic tests and enabling cognitive-based inferences about student performance. We describe the initial approach to developing a cognitive model, the decision-making processes during subsequent iterations of development, and validation of the final cognitive models by content specialists. The strengths and limitations of this method are also discussed and evaluated.

Acknowledgements

The research reported in this study was conducted, in part, with funds provided to the fourth author by the College Entrance Examination Board. We would like to thank the College Entrance Examination Board for their support. However, the authors are solely responsible for the ideas, methods, procedures, and interpretations expressed in this study. Our views do not necessarily reflect those of the College Entrance Examination Board.

Development of Cognitive Models in Mathematics to Promote Diagnostic Inferences about Student Performance

Educational tests should provide meaningful information to guide student learning. The recent emphasis on understanding the psychology underlying test performance has led to developments in cognitive diagnostic assessment (e.g. Leighton & Gierl, 2007a; Mislevy, 2006), which integrates cognitive psychology and educational measurement to enhance learning and instruction. A cognitive diagnostic assessment (CDA) is designed to measure a student's knowledge structures and processing skills. Currently, many educational achievement tests report a small number of content-based subscores, but a CDA reports a profile of scores with specific information about a student's cognitive strengths and weaknesses. This cognitive diagnostic feedback has the potential to guide instructors, parents, and students in their teaching and learning processes.

The development of a CDA begins with articulation of the construct in question by specifying a cognitive model. 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 & Gierl, 2007b, p. 6). These cognitive skills are specified at a small grain size in order to generate specific diagnostic inferences underlying test performance. The resulting cognitive model serves a dual function. It represents the ordered knowledge, skills, and processes students use to solve items associated with measuring the construct. The model also provides a framework for designing diagnostic items for linking examinees' test performance to specific inferences about cognitive skills. The purpose of the assessment, then, is to characterize student performance in terms of the knowledge and skills mastered based on his or her performance on the diagnostic items.

Unfortunately, there are few cognitive models available to use operationally for creating diagnostic assessments because little is known about how students actually solve items on educational tests. The transition necessitated by the use of cognitive models from the achievement to the diagnostic testing domain has been poorly documented and described in the educational and psychological measurement literature. Therefore, the purpose of this paper is to present a method used for developing a cognitive model in Mathematics to promote diagnostic inferences about student performance. This paper describes the development of the initial cognitive model, the decision-making processes during subsequent iterations, and the revisions to the cognitive model using judgments of content specialists. These steps are illustrated in the context of enhancing the diagnostic scoring and reporting components of the PSAT.

Context for the Study

This project ran from Fall of 2004 to the Summer of 2008. The project was funded by the College Board to research cognitive models for college readiness skills. Data from the SAT Reasoning Test and the Preliminary SAT®/National Merit Scholarship Qualifying Test (PSAT) were used. The PSAT is a co-sponsored program by the College Board and National Merit Scholarship Corporation. The PSAT is a standardized test that provides students with practice for the SAT Reasoning Test. It also allows students to enter National Merit Scholarship Corporation scholarship programs. The PSAT measures critical reading skills, math problem-solving skills, and writing skills. The purpose of the PSAT research was to investigate enhanced diagnostic scoring and reporting procedures so that students would receive more specific information about their strengths and weaknesses on college readiness skills. This enhanced feedback was intended to help students focus their preparation on areas where they wanted to improve their test performance. For the PSAT, we developed cognitive models for Critical Reading, Mathematics, and Writing. However, only the results for Mathematics will be described in this manuscript. Cognitive models were created for four content areas in Mathematics: (a) Numbers and Operations, (b) Algebra, (c) Geometry and Measurement, and (d) Data, Statistics, and Probability.

Source of Information

The results of a scale anchoring research study (College Board, 2007) provided the knowledge and skill descriptions used to create an initial draft of the cognitive models in Mathematics. To provide guidance for the hierarchical ordering of these skills, skill categories were taken from O'Callaghan, Morley, and Schwartz (2004). This comprehensive list of the knowledge and skills required to solve mathematics items on the SAT served as the starting point for the PSAT cognitive model development. Any skills that were deemed important but missing in the cognitive models were added from the College Board Standards for College Success: Mathematics & Statistics (College Board, 2006), and updated results from a second SAT scale anchoring research study (College Board, 2008).

Procedure

Overview

Creation of the Mathematics cognitive models proceeded in three stages. The first stage involved development of the initial cognitive models using the knowledge and skills descriptors from an SAT scale anchoring research study (College Board, 2007) and the cognitive skill categories from O'Callaghan et al. (2004). The second stage involved refinement of the cognitive models using the judgments of content specialists in the domain of Mathematics by consensus.

The third stage involved further refinement and finalization of the cognitive models by content specialists reviewing existing PSAT items from the 2005 and 2006 administrations. Throughout the modeling process, the individual skill descriptors were evaluated for grain size, measurability, and instructional relevance.

Stage One: Development of the Preliminary Cognitive Models

Preliminary cognitive models were created to facilitate discussions among the content specialists when creating their cognitive models. The PSAT is usually taken in preparation for the SAT, therefore using the knowledge and skills descriptors from the SAT scale anchoring research study for the PSAT cognitive models maintains the logical connection between the two tests. To begin, O'Callaghan et al.'s (2004) five cognitive skill categories (listed in increasing sophistication) of Applying Mathematical Knowledge 1 (AMK1), Applying Mathematical Knowledge 2 (AMK2), Reasoning (REAS), Managing Complexity (MC), and Creating Representation/Insight (CR/I) were used as the initial framework for creating the cognitive models. A description of these skill categories are provided in Appendix A. The cognitive skill categories represented a developmental progression of mathematical skill development and were deemed an excellent starting point to begin the task of ordering the SAT skill descriptors in terms of increasing complexity.

Initially, one cognitive model was conceptualized for each of the five cognitive skill categories across the four content areas in Mathematics represented on the PSAT. These areas include Number and Operations, Algebra and Functions, Geometry and Measurement, and Data, Analysis, and Probability. SAT skill descriptors for each content area were first coded to their respective cognitive skill category, then related skills were grouped together to create four or five general skills that could be ordered in increasing cognitive difficulty. The skills were ordered based on a judgemental and logical analysis. This process resulted in five cognitive models, one for each cognitive skill category (i.e., AMK1, AMK2, REAS, MC, and CR/I).

The skills identified for a diagnostic assessment should be of a fine grain size in order to make specific inferences about performance in terms of the knowledge and skills used. The grain size of the cognitive skills, initially, in the five cognitive models was too coarse to be diagnostic. Therefore, the grain size of the cognitive skills was refined further and included in the cognitive models for each content area. Again, the skill descriptors for each content area were linearly ordered in increasing cognitive complexity using O'Callaghan's cognitive categories. The skills were ordered based on two considerations: (1) documentation about mathematics curriculum as described in the College Board Standards for College Success: Mathematics & Statistics (College Board, 2006), and (2) the scale score metric of the SAT scale anchoring research study. This process resulted in 20 cognitive models (i.e., five cognitive skill categories by four content areas).

The authors met regularly to provide feedback on each draft of the cognitive models. Comments and suggestions, where appropriate, were incorporated into subsequent revisions.

Stage Two: Revision of the Cognitive Models Using Judgments of Content Specialists by Consensus

Three content specialists in Mathematics, who were involved in the SAT scale anchoring research study (2007, 2008) and had extensive experience working with the SAT and PSAT, were nominated by the College Board to review and revise the proposed cognitive models with respect to the skills and the hierarchical structure. Specifically, their task was threefold: (1) to evaluate the grain size of the skills in the preliminary cognitive models, (2) to evaluate the linear hierarchical ordering of the skill descriptors from least to most complex, and (3) to identify any important skills not included within the cognitive models.

The preliminary cognitive models were reviewed for each content area. Skills in the original cognitive models that were either deleted or re-worded in the updated SAT Performance Category Descriptors (College Board, 2008) were updated in the revised models. Any changes to the cognitive models were agreed upon by mutual consensus among the content specialists.

Stage Three: Revision of the Cognitive Models Using Information from Test Items

The three content specialists were asked to review PSAT items from the 2005 and 2006 administrations and to align these items to the skills identified in the cognitive models. This activity allowed the content specialists an opportunity to revise their models based on operational test items. In particular, this type of information could be helpful to the content specialists for assessing the grain size, measurability, and instructional relevance of the skills which may lead to further revisions to the cognitive models.

Results

Stage One: Development of the Initial Cognitive Model

Twenty cognitive models were created from Stage One of this study. As an illustration, five Stage One cognitive models and their associated skill descriptors in the content area of Algebra and Functions are shown in Figure 1. Linear models were created within each of the five cognitive skill categories of AMK1, AMK2, REAS, MC, and CR/I. Within each of these cognitive skill categories, the skill descriptors were ordered from simple (bottom) to most complex or difficult (top) and missing skills were flagged in red for any gaps in the hierarchy. For example, the AMK1 cognitive model contains five skills, where “Use letters to represent variables in very simple expressions” represents the least cognitively complex skill, and “Verify the solution to a linear equation (e.g., substitute and simplify)" represents the most cognitively complex skill of this hierarchy. A gap in the hierarchy was identified between skill 2.a.3 "Translate a verbal statement into an algebraic equation" and skill 2.a.5 "Verify the solution to a linear equation (e.g., substitute and simplify)". Skill 2.a.4 "Solve one variable linear equation" was flagged as a missing skill required to create a conceptual bridge between skills 2.a.3 and 2.a.5. Another feature of the preliminary cognitive model was the developmental progression or ordering *across* the cognitive skill categories, where AMK1 represented the least cognitively complex category and CR/I the most cognitively complex.

Stage Two: Revision of Cognitive Models Using Judgments of Content Specialists by Consensus

The content specialists had specific criticisms of the original cognitive models which necessitated revisions. Three criticisms were identified: (a) the lack of *process skills* used from the SAT scale anchoring research study (i.e., problem solving, representation, reasoning, connections, and communication) in the cognitive models, (b) the cognitive categories of AMK1, AMK2, REAS, MC, and CR/I were misleading and not representative of the skills classified in the group, and (c) the ordering of the skills in the preliminary cognitive models corresponded to skills increasing in sophistication, but did not always maintain a dependent, hierarchical ordering. Therefore, each cognitive category was renamed Category A to Category E, where the categories were conceptualized as representing a developmental progression of cognitive skill.

During revisions to the models, the content specialists attempted to maintain the structure of the preliminary cognitive models by also using five grouping categories (i.e. Category A to E) for each content area. When revising the hierarchical structure of the skills, the content specialists adjusted the cognitive models to reflect the different connections among the content and skills within the domain of Mathematics. Acquisition of these "connected skills" represented acquisition of proportional reasoning skill, which is considered to be important for successful problem solving in Mathematics. Modeling this connection was accomplished by reusing a number of skills for hierarchies within a content area, and for hierarchies across content areas.

The revisions resulted in an articulation of the skills that, when considered together, are best described as a skill by category matrix (i.e., *n* skills X *5* categories) where skills were ordered from bottom to top in increasing sophistication. An observation was made that conceptual groups which captured the dependency among the skills, were represented by the *rows* of this matrix. That is, the cognitive models could be extracted by *transposing* this matrix. As a result, the maximum length possible for each cognitive model is five, which corresponded to the number of cognitive categories used in the preliminary construct maps. Each cognitive model was renamed and numbered (e.g., Hierarchy 1) for each of the four content areas.

In contrast with the outcomes of Stage One, 38 cognitive models were generated from Stage Two. The increase in the number of models was due to the increase in the number of skills added to the content areas of Geometry and Measurement and Data, Statistics, and Probability. In addition to the increased number of cognitive models between Stage One and Stage Two, there were three important differences: (1) skills were repeated across hierarchies within a content area, (2) the number of skills for each hierarchy was five or less, and (3) no descriptive category labels were provided for the hierarchies. Figure 2 illustrates these changes between the original and revised models in Algebra and Functions, showing 4 out of 13 revised models created by the content specialists. As with the preliminary cognitive models, the skill descriptors were ordered from simple (bottom) to most complex or difficult (top).

Repetition of skills across multiple, linear cognitive models was used to represent the connections between one skill as being a prerequisite to other multiple skills. Referring to Figure

2 for example, the skill "Solve problems involving rates or unit conversions" is considered the least cognitively complex skill for Hierarchies 1, 2, and 3. Additionally, Hierarchies 1 to 4 contained less skills than the original cognitive models (e.g., 3 skills in the revised models compared with 7 skills in the original model). This was due to the process of generating the hierarchical and dependent ordering of skills *across* the five proposed developmental cognitive categories A to E. Using this method, the maximum number of skills found within each hierarchy was five. Although each cognitive model was conceptually coherent, no descriptive category labels were provided at this meeting and would be added at a later date. Again, the focus of this stage was to ensure that the knowledge and skills underlying performance on the PSAT were identified and ordered accurately within the cognitive models.

Stage Three: Revision of the Cognitive Models Using Information from Test Items

At a follow up meeting, the content specialists were asked to align PSAT items from the 2005 and 2006 administrations to their cognitive models created from Stage Two. This activity resulted in minor revisions to the models, providing some evidence that reviewing test items may be helpful to the content specialists when making judgments about the grain size, measurability, and instructional relevance of the skills. In particular, the content specialists paid close attention to the wording of the skill descriptions ensuring that they would be clear and meaningful for reporting purposes, and that they reflected a consistent grain size within the cognitive models. Twelve of the 38 models were changed by adding or removing skills and/or changing the ordering of the skills. In addition, 23 of the 134 (17%) initial skills were modified by either adding or removing words, phrases, and/or descriptions. This process resulted in the creation of two more models, bringing the final total to 40 cognitive models.

Thirteen cognitive models in Algebra and Functions were created from Stage Two of this study. After reviewing the test items, the content specialists created an additional model containing skills they deemed important but was not represented in the second iteration of the cognitive models. An example of four out of the 14 final cognitive models in Algebra and Functions are presented in Figure 3. An example of the small changes that occurred can be seen in Hierarchy 1 from the Stage Two version to the final version. Note that for Hierarchy 1, the content specialists deleted the skill "Solve problems using nonlinear functions" in the final version of the cognitive model because it did not fit in the hierarchy and was considered to be an advanced skill taught at the college level. Also, the content specialists changed the wording in some of the skill descriptors from "and" to "or". The following example illustrates this change, "Interpret and solve word problems using multi-step proportional reasoning involving ratios, rates, **and** proportions" becomes "Interpret and solve word problems using multi-step proportional reasoning involving ratios, rates, **or** proportions" in the final version. The change from "and" to "or" signifies the equivalence of the three mathematical concepts of ratios, rates, and proportions, as well as affording flexibility to the item developers when creating the item to measure this skill in the cognitive model.

Summary and Discussion

The cognitive model is the fundamental basis for developing diagnostic tests and enabling cognitive-based inferences about student performance. The purpose of this paper was to provide a summary of one method for developing cognitive models in the domain of Mathematics for enhancing the diagnostic scoring and reporting components of the PSAT. First, we constructed initial cognitive models based on existing documents outlining the knowledge and skills required to solve items from the SAT. Cognitive models were then constructed for each content area, where the knowledge and skills were ordered in increasing cognitive complexity based on O'Callaghan's hierarchical cognitive categories, as well as the score scale of the SAT. Content specialists in Mathematics were asked to revise the preliminary cognitive models and to evaluate the skills in terms of grain size, measurability, and instructional relevance. The first set of revisions relied primarily on judgments by consensus on the addition, deletion, or modification of the skills and their ordering within the construct map. The second set of revisions used PSAT items from two test administrations as a source of information for modifying the skills within the models. The final cognitive models are a comprehensive representation of the structure of the mathematical knowledge and skills students use to solve the Mathematics section of the PSAT.

Development of the cognitive model is the first step to creating cognitive diagnostic assessments. This form of test development is construct centered, where test items are developed to measure the specific knowledge and skills in the cognitive model. In this way, student performance can then be characterized in terms of the knowledge and skills mastered. This approach to describing the knowledge state of a student differs from that of Falmagne, Koppen, Villano, Doignon, and Johanneson (1990), where knowledge states are described in terms of the types of questions answered correctly. The knowledge and skills required are abstracted from the set of test items answered correctly and are not explicitly grounded in the construct of interest.

Developing the cognitive model as a representation of the construct poses unique challenges. The first challenge lies in identifying the knowledge and skills that characterize performance in a given construct. We used existing performance descriptors and skills from an SAT scale anchoring research study. Another option would be to use knowledge and skills outlined in a curricular document, which students are required to learn during their studies. The second challenge lies in ordering the cognitive skills along a continuum of increasing complexity. One content specialist remarked that there is no metric for ordering these cognitive skills. Yet, intuition would suggest that there must be some developmental progression when learning mathematical concepts. In this study, we relied on the extensive experiences and judgments of the content specialists to order the cognitive skills. Other options include investigations of theory in mathematical cognition and the use of task analyses. The third challenge lies in ensuring that the knowledge and skills specified in the cognitive model are indeed measurable. This last point reminds us that there is a balancing act between the skills we deem important to learn and the skills we are able to operationalize using our current testing formats.

Conclusions

Research efforts in cognitive diagnostic assessment (CDA) have been fuelled by the increasing demand, from both educational stakeholders and researchers, for more formative information from educational tests (Huff & Goodman, 2007). The results of this study can provide guidance to test developers and educational measurement researchers for creating a cognitive model in an achievement domain, which is a crucial step when developing CDAs. In particular, this study demonstrates that creating cognitive models for diagnostic purposes is an iterative process and benefits from the expertise afforded by content specialists. To date, the use of cognitive models in the diagnostic testing domain has been poorly documented and described in the educational and psychological measurement literature. Yet, accurate specification of a cognitive model is vital for creating CDAs that promote diagnostic inferences about student performance. This feedback can then be used by teachers, parents, and students for informing instruction and improving student learning.

References

College Board. (2006). College Board Standards for College Success: Mathematics & Statistics, New York, NY: The College Board.

College Board. (2007). *Performance category descriptions for the critical reading, mathematics, and writing sections of the SAT* (College Board unpublished internal memo). New York, NY: The College Board.

College Board. (2008). *Performance category descriptions for the critical reading, mathematics, and writing sections of the SAT* (College Board unpublished internal memo). New York, NY: The College Board.

Falmagne, J-C., Koppen, M., Villano, M., Doignon, J-P., & Johanneson, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review, 97,* 201-224.

Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 19-60). Cambridge, UK: Cambridge University Press.

Leighton, J. P., & Gierl, M. J. (Eds.). (2007a). *Cognitive diagnostic assessment for education: Theory and applications.* Cambridge, UK: Cambridge University Press.

Leighton, J. P., & Gierl, M. J. (2007b). 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.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257-305). Washington, DC: American Council on Education.

O'Callaghan, R.K., Morley, M.E., & Schwartz, A. (2004, April). *Developing skill categories for the SAT Math section.* Paper presented at the meeting of the National Council on Measurement in Education, San Diego, CA.

Appendix A

Below is a modified description of the cognitive categories by O'Callaghan, Morley, & Schwartz (2004) used in Stage one to hierarchically order the knowledge and skills within the cognitive models for each content area.

Cognitive Category	**Description**
Applying Mathematical Knowledge 1 (AMK1) (concrete, routine problem solving)	Skill involves application of mathematics content and processes that are usually learned in Algebra I or before.
Applying Mathematical Knowledge 2 (AMK2) (less concrete, routine problem solving)	Skill involves application of mathematics content and processes taught in high school Geometry and Algebra II. It also includes more advanced or complex applications of knowledge from Algebra I and before.
Reasoning (REAS) (abstract, less routine problem solving)	If the statistics indicate that a question is difficult, and if that difficulty cannot be totally attributed to the content knowledge required by the question, it is likely that a higher level of reasoning accounts for the difficulty in the question. Questions primarily involving if-then reasoning or the consideration of cases are classified as REAS. Other questions classified as REAS are those that are unfamiliar and have a degree of novelty.
Managing Complexity (MC) (more abstract, non-routine problem solving)	These questions may involve many steps to keep track of or a great deal of information to process. Some questions that are classified as MC are those in which the student must have the "courage" to continue on for each step in the path to the solution. In some cases, that courage may be required to take the first step of plunging in and attempting the solution of a non-routine problem.

Cognitive Category	Description
Creating Representations/Insight (CR/I) (most abstract, very non-routine problem solving)	Questions involving insight are classified as CR/I. Insights are those realizations that often seem easy when someone else points them out but are difficult to see by oneself. In addition, this skill involves modeling or "creating representations". Students with this skill can create equations for word problems that involve more than rote translation. They can also create graphs to solve problems or draw the "insightful" line that makes the solution to a problem become clear. Part of the skill of CR/I is deciding what kind of representation would be useful for solving a problem.

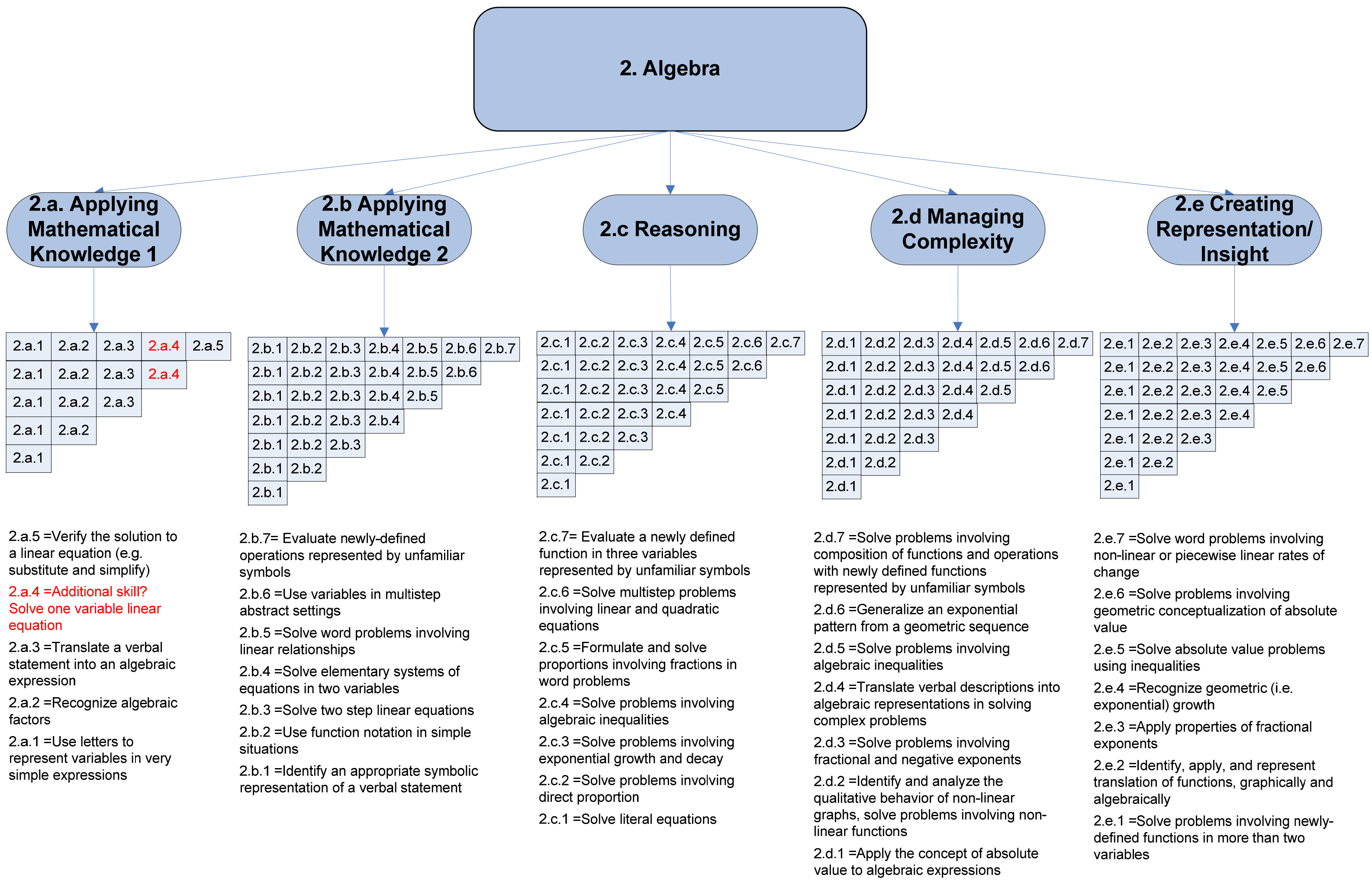
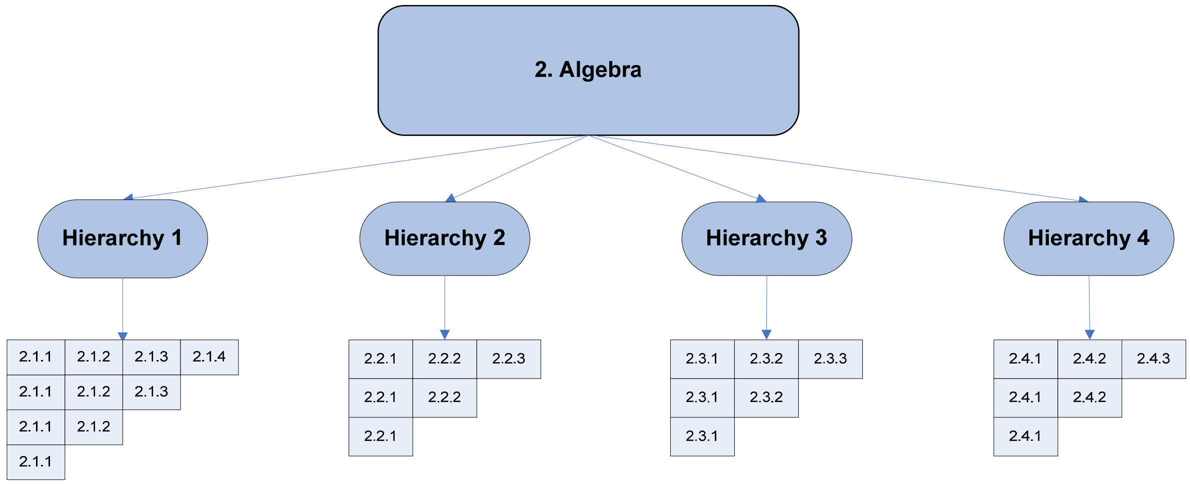


Figure 1. Original proposed construct maps for the PSAT/NMSQT Algebra section



2.1.4 =Solve word problems involving nonlinear or piecewise linear rates of change

2.1.3 =Solve problems involving nonlinear functions

2.1.2 =Interpret and solve word problems using multi-step proportional reasoning involving ratios, rates, and proportions

2.1.1 =Solve problems involving rates and unit conversions

2.2.3 =Identify and use the relationship between the slope of a line and algebraic rate of change

2.2.2 =Use and interpret complex graphs, such as graphs of step functions

2.2.1 =Solve problems involving rates and unit conversions

2.3.3 =Identify and use the relationship between the slope of a line and algebraic rate of change

2.3.2 =Interpret and solve word problems using multi-step proportional reasoning involving ratios, rates and proportions

2.3.1 =Solve problems involving rates and unit conversions

2.4.3 =Solve for one variable or expression in terms of another [Solve literal equations]

2.4.2 =Use variables in multi-step abstract settings (e.g., applying the distributive property across several variables)

2.4.1 =Use a letter as a placeholder for an unknown value

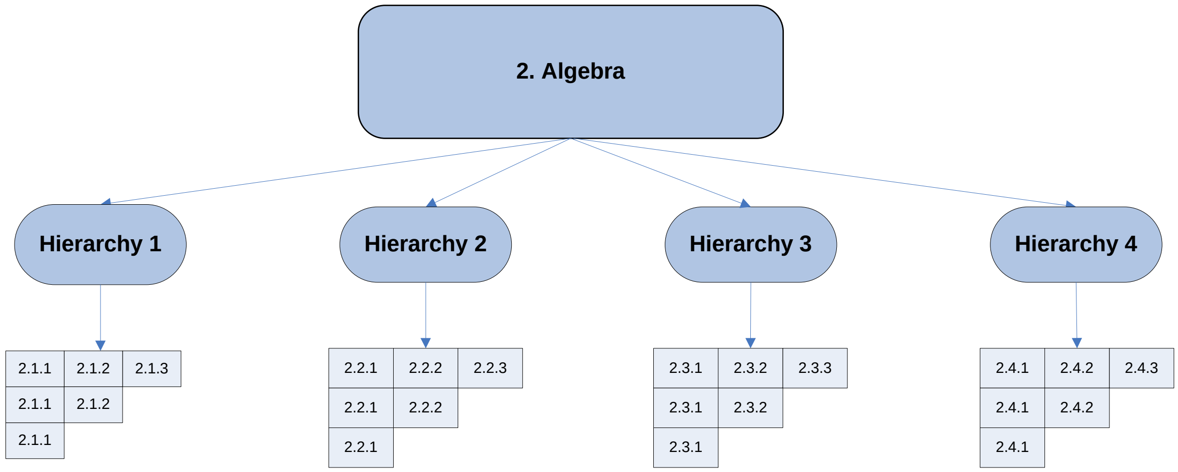
Color Coding for Graphical Construct Map

Red denotes skill that is repeated more than once across hierarchies within one page

Blue denotes skill that is repeated more than once across hierarchies across pages

Purple denotes skill that belongs in more than one content category (e.g., Numbers and Operations and Algebra and Functions)

Figure 2. Four revised construct maps from stage two for the PSAT/NMSQT Algebra section.



2.1.4 =Solve word problems involving piecewise linear rates of change

2.1.2 =Interpret and solve word problems using multi-step proportional reasoning involving ratios, rates, or proportions

2.1.1 =Solve problems involving rates or unit conversions

2.2.3 =Identify and use the relationship between the slope of a line and algebraic rate of change

2.2.2 =Use and interpret complex graphs, such as graphs of step functions

2.2.1 =Solve problems involving rates or unit conversions

2.3.3 =Identify and use the relationship between the slope of a line and algebraic rate of change

2.3.2 =Interpret and solve word problems using multi-step proportional reasoning involving ratios, rates or proportions

2.3.1 =Solve problems involving rates or unit conversions

2.4.3 =Solve for one variable or expression in terms of another [Solve literal equations]

2.4.2 =Use variables in multi-step abstract settings (e.g., applying the distributive property across several variables)

2.4.1 =Use a letter as a placeholder for an unknown value

Color Coding for Graphical Construct Map

Red denotes skill that is repeated more than once across hierarchies within one page

Blue denotes skill that is repeated more than once across hierarchies across pages

Purple denotes skill that belongs in more than one content category (e.g., Numbers and Operations and Algebra and Functions)

Figure 3. Four final construct maps for the PSAT/NMSQT Algebra section.