

Designing Cognitive-Developmental Assessments:
A Case Study in Proportional Reasoning

Gail P. Baxter
Educational Testing Service

Brian Junker
Carnegie Mellon University

Saturday, April 07, 2001

Paper presented at the annual meeting of the National Council for Measurement in Education, Seattle, Washington, April 2001. This work was carried out as part of the NSF-supported New Standards Mathematics Task Bank project, Lauren Resnick and Phil Daro, Principal Investigators (ESI-9876538). Opinions expressed are those of the authors and not necessarily NSF or New Standards. Please do not cite or quote without permission of the authors.

Designing Cognitive-Developmental Assessments:
A Case Study in Proportional Reasoning

Linking testing and teaching in ways that promote high levels of understanding and achievement is a fundamental and enduring concern in education. Yet, despite more than a decade of efforts to redesign curricula, refocus instruction, and promote new conceptions of what it means to know and learn, questions about the quality of education persist. An important tool in answering these questions is an assessment system that is sensitive to changes in subject-matter competence as students proceed through a program of instruction. The pivotal intellectual and practical challenge in designing this assessment system is establishing a cognitive-developmental map in which age-appropriate skills and knowledge competencies can be specified, along with the kinds of performances that can and should be observed under particular circumstances. Explicit descriptions of the development of understanding of a topic can be used to monitor progress toward targeted learning goals, provide visible examples of developing competence, make apparent the gaps in understanding, and suggest activities for promoting further development. To this end, we integrate research on cognition and development with advances in psychometric technique to produce a mathematics assessment system that is diagnostic of programmatic strengths in terms of relevant aspects of student knowledge and performance (Pellegrino, Baxter, Glaser, 1999; Shepard, 2000).

Our work builds on the work of others in mathematics (e.g., Carpenter & Fennema, 1999; Case, 1996) and science (e.g., Hunt & Minstrell, 1994; Minstrell, 2000; Wilson & Sloane, 2000) and is similar in broad concept to the "Progress Maps" developed by ACER in Australia. For example, Masters and Forster (1996) write that, "An essential feature of a progress map is that it describes developing competence in words and examples. A progress map describes the knowledge, skills and understandings of a learning area in the sequence in which they typically develop and provides examples of the kinds of performances and student work typically observed at particular levels of attainment" (pg. 4). Despite the general similarity to the approach

described here, our work must be successful in the more heterogeneous environment of U.S. education where curricular goals and foci vary considerably. In particular we aim to report on student performance variation with respect to specific, and necessarily multidimensional aspects, not unlike the DIAGNOSER project (cf. Minstrell, 2000) rather than with respect to unidimensional task difficulty as induced by the order of exposure to topics and skills in a particular (and shared) curriculum.

Design Considerations. We have begun to explore the design of cognitive-developmental assessments in mathematics. We conceive of this assessment system as a set of four interrelated components:

- (a) A cognitive developmental map that displays the key ideas/concepts and what the literature and teacher-experience say about development of these ideas/concepts during the middle school years;
- (b) A set of assessment tasks that elicit knowledge-based variations in student performance such as flexibility of strategy use, quality of explanation, or nature of problem representation;
- (c) Score criteria that reflect these knowledge-based variations in student performance; and
- (d) A summary of performance so teachers can see how students' performances are related to the cognitive developmental map within and across grade levels.

In selecting topics for a middle school mathematics assessment, we chose to focus on a set of thinking and reasoning skills that: (a) are not dependent on any one curricula but rather they appear (more or less explicitly) in all exemplary middle school mathematics curricula (cf. U.S. Department of Education, 1999); (b) have some prominence in disciplinary standards in mathematics (National Council for Teachers of Mathematics, 1989; 2000) and science (National Research Council, 1996); and (c) have been identified as central to scientific literacy and understanding (cf. American Association for the Advancement of Science, 1993).

In doing so, we can keep cognitive and statistical tasks tractable while at the same time designing an assessment system that is of maximal formative use to teachers. In this paper, we discuss the design considerations involved in constructing a cognitive-developmental map for proportional reasoning.¹

Constructing a Cognitive-Developmental Map

A cognitive-developmental map is a succinct depiction of the performance expectations of developing competence. The "map" when complete, suggests the kinds of tasks or test situations that will elicit knowledge-based differences in student's performance, and the considerations for scoring and reporting methods that capture and reflect relevant variations in the quality of observed performance (cf. Messick, 1994). Thus the formulation of the map is critical to the entire assessment design process. Once drafted, the map undergoes reviews by a group of experts selected specifically for their knowledge/experience with the topic of study, disciplinary standards or curricula, and by a number of selected teachers either individually or in focus groups. The purpose of these reviews is to examine the map and check for consistencies or inconsistencies between the progression of understanding as represented by the map and the progression suggested by those most knowledgeable in the field.

To begin the map design process, we convened a heterogeneous group of scholars and practitioners for a succession of one or two-day meetings over a six-month period. This multidisciplinary team reviewed literature (developmental, cognitive, and mathematics education), collected and analyzed "exemplary" curricula, and considered existing items/tasks designed (in various contexts) to test students' understanding of proportion and ratio. Our goals, broadly speaking, were to describe, to the extent possible, the sequence in which students' knowledge, skills and understandings typically develop and to characterize the nature and patterns of students' performances

¹ Discussion of items/tasks, score criteria, and score summaries awaits data from field trials.

that can be associated with particular levels of attainment (cf. Carpenter & Fennema, 1999; Case, 1996; Masters & Forster, 1996).

Four pieces of information are necessary if we are to realize this goal: (1) a working definition of proportional reasoning, (2) a list of the typical problem types and contexts in which students are required to reason proportionally, (3) a delineation of the aspects of students' performances that constitute evidence of understanding of proportional reasoning, and (4) a description of developmental changes that have been empirically documented. We describe each in turn with attention to the design challenges and implications of each for assessment design.

Definition of Proportional Reasoning

Proportional reasoning is a psychological construct, not a mathematical one. "The essential characteristic of proportional reasoning is that it focuses on describing, predicting or evaluating the relationship between two relationships (i.e., a second-order relationship) rather than simply a relationship between two concrete objects (or two directly perceivable quantities)" (Piaget & Inhelder, 1975 as cited in Lesh, Post, & Behr, 1992, pg. 94). This is an important and difficult conceptual leap for students; mathematical experiences in elementary school focus primarily on countable objects and first-order relationships. In proportional situations students must replace additive reasoning and notions of change in an absolute sense with multiplicative reasoning and notions of change in a relative sense.

In defining proportional reasoning for assessment purposes, we consider three perspectives: psychological, mathematical, and curricular/situational (i.e., the kinds of situations or contexts generated by curricula [for students] or life experiences [for adults] where the problem solver might see proportion and ratio). From a psychological perspective, Vergnaud (1983) has suggested that there is essentially one situational model involved in the understanding of simple proportional relationships, which he terms the isomorphism of measures model of situations. In the isomorphism of measures model, each variable remains independent of the other; parallel

transformations are carried out within or between variables thereby maintaining their values proportional.

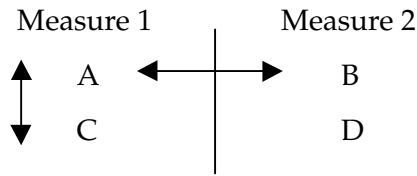


Figure 1. Vergnaud model for simple proportions (1983)

Mathematically, there is one general model ($y = kx$) that could describe the relationship between any two variables linearly related, provided that the y -intercept equals zero. It has no implied directionality: given any three values, y , x and y' , the corresponding x' value can always be computed. Although all proportional reasoning situations can formally be reduced to $y=kx$, students are seldom aware of this.

Students' curricular experiences expose them to two related mathematical models during the middle school years, first equal ratios, expressed as $a/b = c/d$ or $a:b = c:d$, and later linear scaling $y=kx$. Typically, each model is presented (independent of the other) as an algorithm for solving some set of problems. Perhaps due to the sequence in which these models are presented, beginning students see mixing, sharing, and rate comparison problems, in which computing and comparing ratios is appropriate, as quite different from shrinking/stretching problems and those involving familiar intensive quantities (e.g., miles per hour) in which computing with $y=kx$ may be appropriate.

Problem Types and Contexts

There are two major problem types: missing value ($a/b = c/d$) and comparison (which is the better buy?). Continuous scaling problems in which direct use of $y=kx$ would be natural are not encountered until late in most middle school curricula. Here we focus on missing value problems. All missing value problems have the same general format. Students are given three of four values (say a , b , and c), two of the

values are presented as a ratio (a/b) and the task is to determine the missing value (in this case d) such that the ratio of $c/d = a/b$.

Experience with proportion and ratio topics often provide students with their first formal introduction to constants and variables, covariation, invariance, solving for unknowns, and modeling apparently disparate real-world situations with the same linear functions. These experiences begin in some elementary form in the early grades (e.g., sharing pizza) and extend through multiple topics in mathematics and science. For example, proportional reasoning is:

- (a) related to several key areas of the elementary mathematics curricula including, for example, equivalent fractions ($5/3 = n/m$), long division ($805/23=n/1$), place value and percents ($n\% =75/100$), and ratios and rates ($15\text{ft.}/2\text{sec.} = n \text{ miles/hour}$);
- (b) inherently involved with some of the most important algebraic understandings having to do with equivalence, variables, and transformations,
- (c) required for everyday situations such as converting currency (\$1 Canadian = \$.75 U.S.), comparing costs to determine “best buy” (which is the better buy: 12 tickets for \$15 or 20 tickets for \$23), or adjusting recipes for more or fewer people, and
- (d) central to many concepts in science such as density, velocity, and pressure.

Despite its centrality in mathematics and science, students seldom advance beyond rote use of the familiar $a/b = c/d$ algorithm. Indeed U.S. 8th grade students scored below the international average on proportionality on TIMMS (National Center for Education Statistics, 1999). Further, research has shown that student performance is highly sensitive to task and context factors such as type of ratio quantity requested (Karplus, Pulos, & Stage, 1983); the particular numbers in the task (Hart, 1984); and the context of the problem (Noelting, 1980a, b; Vergnaud, 1983). For example, it is generally the case that problems involving common numbers (5, 10, 100, integers, and

familiar fractions (such as $1/2$ and $1/3$ and later $1/4$) are easier than problems involving uncommon numbers, non-integers, and non-familiar fractions or decimals. Finally, we note that in a systematic analysis of missing value problems, Harel and Behr (1989) identified 512 missing-value problem structures! Consistent with Vergnaud (1983), they find that location of the unknown and nature of the ratio (e.g., familiar multiples like 3: 9 or 1:4 easier) influence task difficulty.

Evidence of Understanding

In considering evidence of student's understanding of proportional reasoning, three broad categories appear relevant: moderating effects such as arithmetic difficulty, or cover story for the task (i.e., context), solution strategies, and conceptual underpinnings. Context was discussed in the previous section. Here we focus on solution strategies and conceptual underpinnings.

Strategy Use. A review of the literature suggests that a key distinguishing feature of performance is strategy use. Recall, Vergnaud (1983, 1994) characterizes simple direct proportions as isomorphism of measures problems that include situations (or contexts) such as equal sharing, constant price, or uniform speed. He elaborates on four subclasses of problems as formulated in table 1 below. The particular formulation has implications for student strategy use and probability of getting the correct answer (see below).

We point out four features of these general categories of problems. First, categories 1, 2, and 3 all provide a unit value $f(1)$ which facilitates problem solution. Second, category 1 and 4 problems require multiplication (scaling up) and category 2 and 3 problems require division (scaling down). Scaling up (e.g., if one doll costs \$5, how much do 3 cost?) is easier than scaling down (e.g., if 3 dolls cost \$15, how much does one doll cost?).

Third, categories 1 and 2 present a simple ratio or comparison (one involving a unit rate, 1 as part of the ratio) *within* a single measure space or variable and categories 1 and 3 present a ratio or comparison *between* measure spaces. The simple ratio suggests a scalar solution strategy that focuses upon changes within a measure space or

variable. In such approaches a person compares, for example, how many items were bought in the first instance (for a given price) with the number of items to be bought in the second instance. The essential idea is that the relative increase (or decrease) in goods cost should be directly related to the increase (or decrease) in goods purchased.

Table 1. Isomorphism of Measures Problems (from Vergnaud 1983).

Missing Value Problems							
1		2		3		4	
Given $f(1)$ and x , find $f(x)$		Given x and $f(x)$, find $f(1)$.		Given $f(x)$ and $f(1)$, find x		General case: none of the values is set to 1.	
Richard buys 4 cakes at 15c each, how much does he have to pay?		Susan 's mother gave her 12 sweets to share with her friends Barbie and Ken. How many sweets will each get?		Peter has \$15 to spend and he would like to buy miniature cars. They cost \$3.00 each. How many cars can he buy?		When making jam, my grandmother uses 3.5 kg of sugar for 5 kg of berries. How much sugar does she need for 8 kg. of berries.	
Cakes	Cost	Friends	Sweets	Cars	Cost	Sugar	Berries
A = 1	C = 15	A = 1	C = ??	A = 1	C = 3	A = 3.5	B = 5
B = 4	D = ??	B = 3	D = 12	B = ??	D = 15	C = ??	D = 8
Presented ratio: either (within or between)		Presented ratio: within		Presented ratio: between		Presented ratio: either	
Suggested strategy: either		Suggested strategy: within (scalar)		Suggested strategy: between (functional)		Suggested strategy $a/b = c/d$	
Resulting ratio: either		Resulting ratio: between (unit rate)		Resulting ratio: within (scale factor)		Resulting ratio: either	

In contrast, presenting a simple ratio between measure spaces suggests a functional approach that focuses on relationships between variables or measure spaces. The emphasis here is on how one variable varies as a function of the other variable. There is strong evidence in the literature that “within” comparisons are easier than “between” comparisons, perhaps because they allow for pseudo-scalar strategies such as pattern matching, coordinated build-up and factor of change which are intuitive extensions of students’ early understandings of covariance and invariance (cf. Kaput & West, 1994). These strategies are not typically taught in school but arise naturally in- and out-of-school contexts. For this reason, for example, studies of both in- and out-of-

school proportional reasoning (Hoyles, Noss, & Pozzi, 2001; Schliemann & Nunes, 1990; Schliemann & Carraher, 1993; Vergnaud, 1983) show that problem solvers tend to strongly prefer “within” or “scalar” strategies even when “between” or “functional” ones are easier computationally (as in category 3 above).

Finally, a quick review of the four problem types indicates that each problem can be correctly solved in several different ways (e.g., scalar, functional, build up, unit factor). The correct answer is not an indicator of strategy use and indeed students often choose a strategy not predicted by the problem type (see table 1 above). The diagnostic value of the item can be improved by providing each rule or strategy-based rationale with its corresponding numerical response. This is similar to the DIAGNOSER approach; Students are given feedback on the consistency between their answer and their explanation or strategy use (cf. Hunt & Minstrell, 1994; Minstrell, 2000).

While on the surface this matching of answer to strategy might seem like a suitable approach (particularly for multiple-choice items), it does not address the variation in strategy use across contexts. A number of different studies that vary with respect to age of student, type of problem, and variation in problem contexts (Hart, 1984; Karplus, Pulos, & Stage, 1983; Kaput & West, 1994; Noelting, 1980a,b) have found that students’ strategies are inconsistent across tasks and sometimes students change strategies during a task (e.g., when the number relationships, perceptual distractors, or contextual variables are changed slightly). For many students with some experience in proportional reasoning, their strategy use may appear “expert-like” in familiar situations, but changes in the problem context (e.g., difficult arithmetic generally, or presenting integer values in a between-measure-spaces comparison that are very close together) often results in them opting for an incorrect additive reasoning representation and strategy (cf. Kaput & West, 1994).

Conceptual Underpinnings. Recall that proportional situations require students to reason about the relation between relations. For any given context, students must distinguish between situations in which comparisons are proportional or multiplicative (e.g., money exchange) versus additive (e.g., father’s age to son’s age) and then carry

out a solution strategy such that the component measures of a ratio are varied in such a way that the original ratio relationship remains invariant (i.e., covariance and invariance). Some problems and/or students' solutions (as noted above) do not provide direct information on students' understanding of these key ideas. Direct probes may be necessary if we are to provide diagnostic information to inform instruction (cf. Lamon, 1993a, b; 1994; 1999a and Figure 2).

Students must be able to understand changes in two different perspectives: absolute and relative.	Store A advertises a sale of \$10.00 off. Store B advertises a sale of 10% off. If both stores sell the same things, which store offers the best sale?
Students must be able recognize the invariant nature of relationships under change.	In Mr. Trent's science lab, there are 3 people to each table. For mid-morning break, Mr. Trent put 2 candy bars on each table and told the students to split them fairly. "Before you start your snacks, though," says Mr. Trent, "I would like you to push all four tables together. Would you rather get your share of the candy before the tables are pushed together or after?"
Students must be able to make decisions in real-world contexts.	Suppose you have 5 cookies and three children. If they share fairly, how many cookies does each child get? Now suppose you have 5 books and 3 children. Can they still share fairly?
Students must be able to understand part-whole and part-part comparisons.	In a group consisting of 24 boys and 6 girls, what percentage of the class are girls?

Figure 2. Conceptual underpinnings.

Developmental Sequence. Proportional reasoning has a rich history of research in developmental psychology (e.g., Kaput & West, 1994; Noelting, 1980a, b), mathematics education (Lamon, 1994, 1999a; Lesh, Post, & Behr, 1988), and science education (Karplus, Pulos, & Stage, 1983). From this work we know that proportional reasoning develops over a long period of time and continues to be problematic for students in middle school and beyond (e.g., Hart, 1984; Kaput & West, 1994). For example only 12% of 8th graders could use proportions (or rates) to solve a problem as seen on a recent NAEP survey in mathematics (Wearne, & Kouba, 2000). Despite the difficulty students have with proportional reasoning, numerous studies suggest that progress (albeit slow) can be traced through a set of developmental stages. As proposed, these

stages are distinguished by qualitative changes in the nature of students' responses to proportional reasoning problems (see Figure 3). Key among the changes is a shift from additive to multiplicative reasoning and from use of context-specific strategies to a generalized understanding of functions.

Qualitative. Young students generally possess a good deal of knowledge about quantity that permits them to answer questions about more and less (e.g., which drink is sweeter?) or fairness (e.g., divide pizza or cookies so everyone gets a fair share).

Early Attempts at Quantifying. Early attempts at quantifying often involve constant additive differences (i.e., $a - b = c - d$) rather than multiplicative relationships. Still rely on counting up or down.

Recognition of Multiplicative Relationship. Students have the intuition that the differences change with the size of the numbers and that the change may be multiplicative in nature, but they do not necessarily realize that they need to consider the constantly increasing/decreasing difference between the related terms of each pair, that is, of each ratio. They rely on pattern matching or build-up strategies that are sometimes additive in nature. Problems with unit factors or sharing are easily solved but others may lead to additive solution strategies.

Accommodating Covariance and Invariance. Students begin to develop a change model that recognizes that while some quantities may be changing, relationships among the quantities remain invariant. However, they rely on coordinated build up, unit factor and other scalar approaches. Often strategies are context specific and not generalizable. May revert to additive strategies in unfamiliar or challenging situations.

Functional and Scalar Relationships. Students recognize the invariant nature of the relationships between pairs of changing quantities. Have generalized model for solving proportionality problems. This is not to say that they use the same strategy in every context, rather they have a repertoire of strategies and use the most efficient for a given situation.

Figure 3. Development of proportional reasoning.

As in other domains, "expertise" in proportional reasoning is marked, not by use of a single most sophisticated strategy but, by a familiarity with--and facility moving between--several strategies and problem representations in response to different situational conditions or problem constraints (e.g., Anderson, 1985; Siegler, 1988). It should be noted that the school-taught algorithm may not be the most efficient strategy in some situations. Although it is almost never used outside of school, it can facilitate a solution when the numbers are hard or the context less familiar; Use of this strategy does not necessarily signal understanding of proportions or proportional relationships.

As noted in our discussion of strategies, students have a preference for scalar over functional strategies or the rule-of-three strategy. We believe that part of what makes scalar strategies attractive is that they are easy to remember within the semantics of early encounters with covariance and invariance and early manipulation of fractions. Thus we view the many versions of scalar strategies (skip counting, parallel build up, abbreviated build up (cf. Lamon 1999; Kaput and West, 1996) as early procedural tools that children acquire to solve specific kinds of problems, before their understanding of proportional reasoning is complete enough to allow them to select strategies more or less independently of the semantics of the problem context or cover story. Evidence from studies in out-of-school contexts where dealing with proportional situations is a part of everyday life such as street vendors (computing cost of goods), or construction workers (scaling up from blue prints), and nurses (calculating dosage of medicine) suggests something similar is going on in adults who have not acquired a broad, more or less context independent, understanding of proportional reasoning (cf. Hoyles, Noss, & Pozzi, 2001; Schliemann & Carraher, 1993.)

Thus, we might expect most middle school students to exhibit performances somewhere between recognition of multiplicative relationships and accommodating covariance and invariance. We expect few students in middle school to regularly employ fully scalar or functional strategies, and we also expect few to employ only qualitative understandings. Further within these general stages, we expect to see substantial variation in students' problem solving performance in direct response to variations in aspects of the problem such as context, location of the unknown (as in Vergnaud models above), or difficulty of the numbers.

Some Remarks on Modeling

Although this paper has only touched on a part of the domain of proportional reasoning, from psychological, mathematical and curricular/situational perspectives, it is clear that we can enumerate most or all of the various possible strategies, qualitative understandings, and moderating hurdles (arithmetic, context) that students bring to

proportional reasoning tasks. From the literature reviewed here, we know for example that, (1) observed performance (or perhaps we should say strategy) is dependent on context, (2) a given answer can reflect several different strategies, and (3) particular strategy use is less important than flexibility and efficiency of strategy use across contexts.

These aspects of performance present some significant challenges for choosing number and type of tasks (given practical constraints on testing time), and for scoring (modeling), and reporting. With unlimited testing time and careful elicitation and recording of students' behaviors as evidence, we could make more or less direct inferences about the specific "location" of individual students within the web of skills, strategies and understandings that constitutes proportional reasoning.

However, we do not have infinite testing time, and we may have quite some difficulty eliciting and recording informative behaviors from each student. Thus for each student there will be missing data: not all questions that we might like to ask will be asked, and not all behaviors that would be informative about students can or will be observed. One way to deal with this "missing data" problem is to use latent variable models because they provide a convenient way to share and borrow information across students in a testing situation, in a way that is equivalent to the sharing of information in a hierarchical Bayesian model or a Bayes network.

For example, consider the item-response-theory-like framework for modeling strategy shifts in problem solving processes formally introduced by Rijkens (1996). In the Rijkens framework, each student is modeled in terms of three components: (1) a continuous latent variable representing "level of development" in proportional reasoning, (2) a limited set of "ideal" categories, or types, of responses for each task (representing alternative solution strategies, alternative qualitative understandings, etc., as is appropriate to the task), and (3) a limited number of patterns of observed evidence produced by examinees' task performances (e.g. types of verbal explanations, types of intermediate calculation, etc.) related to the ideal response types.

Among the “ideal” types of responses there may be several that lead to a “correct” final answer and several that lead to an “incorrect” final answer. Thus the “ideal” types are not ideal in the sense that they lead to the right answer, but rather they are ideal in that they are well matched to typical strategy choices or understandings about the task at various levels of development. These “ideal” types are related to the latent variable representing students’ developmental level via a standard latent variable polytomous logistic regression model, such as Bock’s (1972) nominal response model. From this perspective, a high level of development favors this ideal response type; a lower level favors an alternate response type. The ideal types may be repeated for different tasks, and need not be ordered in the same way on different tasks, reflecting the notion discussed above that problem solvers move between strategies and problem representations in response to different situational conditions or problem constraints. The probability of seeing a certain pattern of observable evidence in a student’s task performance varies with the ideal response type (strategy, understanding, etc.) used by the student. Conversely the pattern of evidence can be used to infer which ideal response type was used by the student (in a probabilistic sense).

Thus, the pattern of evidence within a task performance allows us to make inferences about which strategies, representations and understandings the student is employing *within* each task; and the pattern of these strategies, representations and understandings *across* tasks allows us to make inferences, through estimation of the latent variable, about the level of development of proportional reasoning for each student. These inferences can be generalized to groups of students, and progress may be tracked over time, through estimation of the latent variable distribution and subsequent prediction of task performance for a group of students. Recent experience with a mathematically related model (Patz, Junker, & Johnson, 2000) suggests that estimation and inference issues with this model will be tractable.

A potential drawback of this approach is that the continuous latent variable representing level of development may impose developmental ordering constraints that are not matched by student’s actual progress in proportional reasoning. For example,

depending on curricular and out-of-school experiences, it may be that not all children pass through the stages Qualitative, Early Quantitative, Multiplicative, Covariance/Invariance, and Functional/Scalar in exactly the order presented in Figure 3 above. This may be especially true in the middle grades where curricular choices may affect whether one moves, for example, from the early stages to the Multiplicative and then the Covariance/Invariance stages, or from the early stages to Covariance/Invariance, followed by a more formal recognition of Multiplicative relationships.

We may correct this drawback by replacing the continuous level-of-development variable in the first part of Rijkjes' framework with a discrete latent class variable indicating a student's current "location" in one of five developmental stages. The remaining two parts of Rijkjes' framework remain essentially the same. Such an approach allows us to (a) accommodate different developmental paths and avoid assigning an order to the developmental stages in the statistical model; and (b) test hypotheses about the order of the developmental stages, for example by comparing this "unordered" version of the model with a version that imposes ordinal constraints. Similarly, progress of a group of students over time (or cross-sectional comparisons), will "emerge" as the proportion of students in each developmental category changes with various characteristics of the students such as age or grade level.

The modified, latent class version of the Rijkjes' model is in fact a version of the discrete Bayes networks that have proven to be useful in other cognitively-motivated assessment settings (e.g., Martin & VanLehn, 1995). It is also closely related to the fuzzy set approach taken by Moore, Dixon, and Haines (1991) in their analysis of proportional reasoning in certain mixture tasks. An apparent strength of the Rijkjes' modeling framework is that it does not put any constraints on the task format that can be used to generate evidence with which to make inferences about students' understanding. Coded student work on missing value and comparison problems of the type discussed above (see section on problem types and contexts) can be used, as can coded student responses to conceptual questions of the types discussed in the section on evidence of

student understanding. The coding can be done by hand from raw student work (verbal or written explanations, intermediate written work, etc.) or it can be done automatically in an intelligently-designed multiple choice format, as in the DIAGNOSER assessments.

Two things distinguish our approach from many past implementations. First, we have explicit interests in reliably tracking the progress of a group of learners as they move across relatively large, coherent stages of development in a domain; thus our descriptors of the “ingredients” of a task performance are more coarse and allow more flexibility for “falling off” the map (through the probabilistic relationship between ideal response types and observable patterns of performance) than are typical in some other settings (cf. Nichols, Chipman, & Brennan, 1995). Second, both the discreteness and the coarseness of our approach suggest that simple summaries computable by teachers or administrators in classrooms and schools should be closely related to parameter estimates when the full model is fitted. Some exploration of this latter issue is already under way with a related, simpler model (Junker and Sijtsma, 2000; see also Junker 2001.)

Concluding Comment

The challenge in designing a cognitive-developmental map is incorporating what we know about student performance variation and its relationship to student understanding in ways that can guide assessment design, scoring and reporting. We have begun to meet this challenge with one topic in mathematics, proportional reasoning. Here we reported on the definition of proportional reasoning, the problem types and contexts in which they are situated, and some observable features of student performance that are indicative of more or less proficiency (e.g., strategies) as derived from our synthesis of the extant literature. Taken together, these pieces of information provide a set of expectations (or hypotheses) that can guide task design and empirical evaluation in the next phase of our work.

References

- American Association for the Advancement of Science. (1993). Benchmarks for Science Literacy. New York: Oxford University Press.
- Anderson, J. R. (1980). Cognitive psychology and its implications. San Francisco: Freeman.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37(1), 29-51.
- Carpenter, T. P., & Fennema, E. (1999). Children's mathematics. Cognitively-guided instruction. Heinemann.
- Case, R. (1996). Changing views of knowledge and their impact on educational research and practice. In D. R. Olson & N. Torrance (Eds.), The handbook of education and human development: New models of learning, teaching, and schooling, pp. 75-99. Oxford, England: Blackwell Publishers, Inc.
- Glaser, R., & Baxter, G. P. (2000). Cognition and construct validity: Evidence for the nature of cognitive performance in assessment situations. In H. Braun & D. Wiley (Eds.), The role of construct in psychological and educational measurement. A Festschrift in honor of Sam Messick. Hillsdale, NJ: Erlbaum.
- Harel, G., & Behr, M. (1989). Structure and hierarchy of missing-value proportion problems and their representations. Journal of Mathematical Behavior, 8, 77-119.
- Hart, K. (1984). Ratio: Children's strategies and errors. Windsor, England: NFER-Nelson.
- Hoyles, C., Noss, R., & Pozzi, S. (2001). Proportional reasoning in nursing practice. Journal of Research in Mathematics Education, 32 (1), 4-27.
- Hunt, E., & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), Classroom lessons. MIT Press.
- Junker, B. W. (1999). Some statistical models and computational methods that may be useful for cognitively-relevant assessment. Commissioned paper for the Committee on the Foundations of Educational and Psychological Assessment, National Research Council.
- Junker, B. W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. A. Boomsma, M. A. J. Van Duijn, & T. A. B. Snijders (Eds.). Essays on item response theory. New York: Springer-Verlag.
- Junker, B. W. & Sijtsma, K. (2000). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. Working paper.
- Kaput, J. J., & West, M. M. (1994). Missing-value proportional reasoning problems: Factors affecting informal reasoning patterns. In G. Harel & J. Confrey, (eds). The development of multiplicative reasoning in the learning of mathematics, pg. 235-287. Albany NY: SUNY Press.
- Karplus, R., Pulos, S., & Stage, E. K. (1983). Proportional reasoning of early adolescents. In R. Lesh and M. Landau (Eds.), Acquisition of mathematics concepts and processes, pg. 45-86. New York: Academic Press.
- Lamon, S. J. (1993a). Ratio and proportion. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), Rational numbers: An integration of research (pp. 131-156). Hillsdale, NJ: Erlbaum.

- Lamon, S. J. (1993b). Ratio and proportion: Connecting content and children's thinking. Journal for Research in Mathematics Education, 24(1), 41-61.
- Lamon, S. J. (1994). Ratio and proportion: Cognitive foundations in unitizing and norming. In G. Harel & J. Confrey (eds). The development of multiplicative reasoning in the learning of mathematics. Albany NY: SUNY Press.
- Lamon, S. J. (1999a). Teaching fractions and ratios for understanding : Essential content knowledge and instructional strategies for teachers. Mahwah NJ: Lawrence Erlbaum.
- Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), Number concepts and operations in the middle grades (pp. 93-118). Reston, VA: National Council of Teachers of Mathematics.
- Martin, J., & VanLehn, K. (1995). A Bayesian approach to cognitive assessment. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.) Cognitively diagnostic assessment. Hillsdale, NJ: Lawrence Erlbaum Associate.
- Masters, G., & Forster, M. (1996). Progress Maps Assessment Resource Kit. Melbourne, Australia: Australian Council for Educational Research Ltd.
- Messick, S. J. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13-23.
- Minstrell, J. (2000). Student thinking and related instruction: Creating a facet-based learning environment. In J. Pellegrino, L. Jones, & K. Mitchell (Eds.) Grading the Nation's Report Card: Research for the Evaluation of NAEP. Committee on the Evaluation of NAEP, Board on Testing and Assessment. Washington, DC: National Academy Press.
- Mislevy, R. J. (1996). Test theory reconceived. Journal of Educational Measurement, 33, 379-416 .
- Moore, C. F., Dixon, J. A., & Haines, B. A. (1991). Components of understanding of proportional reasoning: A fuzzy set representation of developmental progressions. Child Development, 62, 441-459.
- National Center for Education Statistics (1999). The Third International Mathematics and Science Study. Overview and key findings across grade levels. Washington, DC: Office of Educational Research and Improvement, US Department of Education.
- National Council of Teachers of Mathematics (2000). Principles and standards for school mathematics. Author.
- National Research Council (1996). National science education standards. Washington, D. C.: National Academy Press.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (1995). Cognitively diagnostic assessment. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Noelting, G. (1980). The development of proportional reasoning and the ratio concept: Part I—Differentiation of stages. Educational Studies in Mathematics, 11, 217-253.
- Noelting, G. (1980). The development of proportional reasoning and the ratio concept: Part II—Problem structure at successive stages; problem solving strategies and the mechanism of adaptive restructuring. Educational Studies in Mathematics, 11, 331-363.

- Patz, R. J., Junker, B. W., & Johnson, M. S. (2000). The hierarchical rater model and its application to large-scale educational assessment data. Submitted for publication.
- Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the "Two Disciplines" Problem: Linking theories of cognition and learning with assessment and instructional practice. Review of Research in Education, 24, 307-353.
- Rijkens, C. P. M. (1996). Testing hypotheses on cognitive processes using IRT models. Unpublished Doctoral Dissertation, University of Twente, Enschede, the Netherlands
- Schliemann, A. D., & Carraher, D. W. (1993). Proportional reasoning out of school. In P. Light & G. Butterworth (Eds). Context and cognition: Ways of learning and knowing, pp. 47-73. London: Harvester Wheatsheaf.
- Schliemann, A. D., & Nunes, T. (1990). A situated schema of proportionality. British Journal of Developmental Psychology, 8, 259-268.
- Shepard, L. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 1-14.
- Siegler, R. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. Child Development, 59, 833-851.
- U.S. Department of Education (1999). Exemplary and promising mathematics programs. http://www.enc.org/professional/federalresources/exemplary/promising/documents/0,1341,CDS-000496-496_1.htm
- Vergnaud, G. (1983). Multiplicative structures. In R. Lesh and M. Landau (Eds), Acquisition of mathematics concepts and processes, pp. 127-174. New York: Academic Press.
- Vergnaud, G. (1994). Multiplicative conceptual field: What and why? In G. Harel and J. Confrey (Eds.), The development of multiplicative reasoning in the learning of mathematics, pp. 41-59. Albany, NY: State University of New York Press.
- Wearne, D., & Kouba, V. L. (2000). Rational numbers. In E. A. Silver and P. A. Kenney (Eds). Results of the seventh mathematics assessment of the National Assessment of Educational Progress. Reston, VA: National Council of Teachers of Mathematics.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. Applied Measurement in Education, 13(2), 181-208.