

# Chapter 3

## Rapid Dynamic Assessment for Learning

Slava Kalyuga

### 3.1 Introduction

Enhancing formative diagnostic assessment is a clear current trend in educational testing. Such assessment allows determining specific levels of acquisition of knowledge and skills and provides fine-grained diagnostic information about strengths and weaknesses of a particular learner. Teachers are encouraged to use more formative assessments throughout their courses to inform their classroom instruction.

In a major report on educational assessment, Pellegrino et al. (2001) emphasized that cognitive theories should be the cornerstone of the assessment design process directed toward evaluating students' schematic knowledge structures. Cognitive models of specific domains are usually based on task analyses, expert interviews, and verbal protocols of thinking processes and identify cognitive attributes required for successful learning and performance in these domains. The need for using cognitive theories of learning and models of expertise as foundations for the design of assessment has been recognized by many educational testing theorists (Embretson 1993; Mislevy 1996; Pellegrino et al. 1999; Pirolli and Wilson 1998; Snow and Lohman 1989; Tatsuoka 1990).

Cognitive diagnostic assessment is aimed at providing ongoing information about students' mastery of specific cognitive processes and operations required for learning and performing particular types of tasks. It combines cognitive models of corresponding domains and statistical models of students' response patterns. Empirical evidence shows that cognitive diagnostic assessment is capable of maximizing students' learning outcomes (e.g., Russell et al. 2009).

However, testing learners continuously without interfering with their learning is a challenging task. Testing time could not be increased considerably as it would

---

S. Kalyuga (✉)

School of Education, University of New South Wales, Sydney, NSW, Australia  
e-mail: s.kalyuga@unsw.edu.au

inevitably reduce instruction time. Traditional standardized multiple-choice tests are rather time consuming and not always represent the best way of diagnosing learner actual levels of knowledge in a domain. With most currently used diagnostic assessment techniques, developing and administering the tests, obtaining data, and interpreting results, as well as incorporating appropriate instructional interventions based on these results, may require considerable amount of time. As a consequence, many teachers may not be inclined to use cognitive diagnostic assessment to guide their instructional decisions.

A possible solution for this problem is to make diagnostic assessment rapid in order to accelerate the process (rapid diagnostic assessment). Another possibility is to use diagnostic assessment itself as an instructional means by integrating seamlessly testing and learning. With this approach, students learn while being tested or are assessed while learning (dynamic assessment). This chapter starts with the description of the general idea of a rapid diagnostic assessment approach and its theoretical framework based on cognitive nature of expertise, schema-based assessment, and cognitive load theory. Then, it describes a general design approach and its specific implementations as rapid diagnostic methods, as well as their possible integration with dynamic assessment methods (rapid dynamic assessment). A summary and directions for further research and development in this area conclude the chapter.

## **3.2 Theoretical Framework**

### ***3.2.1 Knowledge Base and the Nature of Expertise***

Whether expertise is considered in a real professional sense (e.g., Ericsson and Charness 1994) or at a narrow task-specific level (e.g., secondary school students as experts in solving linear algebra equations), it includes a well-organized domain-specific knowledge base as its most important component (Bransford et al. 2000). This knowledge resides in long-term memory which represents one of the major components of human cognitive architecture that underlies cognition and learning. Another essential component of this architecture is working memory.

According to a contemporary model of human cognitive architecture (Sweller 2004; Sweller et al. 1998; Van Merriënboer and Sweller 2005), working memory represents our immediate conscious processor of information. It is limited in both duration and capacity when dealing with novel elements of information (Baddeley 1997; Miller 1956; Peterson and Peterson 1959). No more than a few elements of information could be processed and maintained consciously at the same time in working memory, and they would most likely be lost after a few seconds (unless intentionally rehearsed). For a simple example, consider dialing an unfamiliar mobile phone number after just having heard it from another person.

In familiar domains, the available knowledge base in long-term memory allows us to chunk many elements of information in larger units that could be treated as single elements in working memory. For example, it would be easier to dial the above phone number if you notice a well-known combination of digits as a part of it (e.g., “2010” that could be treated as a single unit of information instead of four). Therefore, the long-term memory knowledge base effectively influences the actual content and capacity of working memory and determines the efficiency of performance.

Studies of expert-novice differences in cognitive science have convincingly demonstrated that learner knowledge base is a single most important cognitive characteristic that influences learning and performance (e.g., Chi et al. 1981; Larkin et al. 1980; see Pellegrino et al. 2001, for a review). When experts face a problem in a familiar area, their available knowledge structures are rapidly activated and brought into working memory as problem-relevant chunks of information. Ericsson and Kintsch (1995) called such knowledge structures associated with currently active working memory elements as long-term working memory (LTWM) structure. They are capable of holding virtually unlimited amount of information due to the chunking effect. In the absence of appropriate domain-specific knowledge structures, novices have to resort to cognitively inefficient and time-consuming random search or weak problem-solving methods such as means-ends analysis or trial-and-error approach.

For example, in classical studies of chess expertise by De Groot (1965) and Chase and Simon (1973), professional grand masters performed considerably better than amateur players in reproducing briefly presented chess positions taken from real games, although there were no significant differences when random configurations of chess figures were used. Knowledge of effective moves for a large number of different real game patterns held in grand masters’ long-term memory allowed them to reproduce chess positions by large chunks of familiar patterns rather than by individual chess figures. During a short exposure to a real-game board configuration, they were able to form long-term working memory structures associated with presented configurations of chess figures using their available domain-specific knowledge base.

The organized generic knowledge structures that we use for categorizing information according to familiar patterns are called schemas. Since the levels of learner expertise in a specific domain are determined by the levels of acquisition of schematic knowledge structures in long-term memory, schemas should be the major target for diagnostic assessment of expertise. In cognitive science, laboratory studies using interviews, observations, and “think aloud” protocols are conducted for uncovering schemas held by individuals (Chi et al. 1989; Ericsson and Simon 1993; Magliano and Millis 2003). Although highly powerful and precise, these methods are very time consuming, slow, and not suitable for realistic educational settings. Combining high levels of diagnostic power with acceptable speed of assessment and simplicity of its implementation is a very challenging task. The next section describes an idea of a potentially suitable approach.

### 3.2.2 *Rapid Schema-Based Assessment*

Since long-term memory that contains schematic knowledge base cannot be accessed directly, we usually make inferences about learners' available knowledge structures based on the results of their problem-solving performance (e.g., answers to multiple-choice items or recorded problem solutions). However, such inferences may not be reliable because they are based on remote and indirect results of actual cognitive processes and structures. They could in fact be misleading for cognitive diagnosis.

For example, based on students' correct answers to multiple-choice items in solving algebra equations (e.g.,  $5x = -4$ ), it is not possible to say exactly what solving processes were involved. Some students could apply knowledge-based schematic solution procedures, but others could achieve the same outcomes by using a random search method. Even those students who relied on knowledge structures could use different levels of knowledge. Some students could apply fine-grained step-by-step procedures (dividing both sides of the equation by 5,  $5x/5 = -4/5$ , then canceling the same numbers in the numerator and denominator on the left side of the equation), while others could use higher-level automated procedures by skipping intermediate steps with the final answer ( $x = -4/5$ ) obtained immediately. Traditional multiple-choice tests would place these two groups of students who are at correspondingly intermediate and top levels of expertise in this task area, together with novices using weak problem-solving methods, in the same category of successful learners (Kalyuga 2006b, d).

A similar situation could be with traditional methods used for assessing reading skills that do not measure students' actual cognitive representations constructed during reading (Magliano and Millis 2003). Students are usually required to read segments of text and then answer multiple-choice questions related to the concurrently displayed texts. Correct answers to such multiple-choice questions would not indicate what actual cognitive processes were used before selecting those answers. Students who achieved correct answers by repeatedly searching the text for key question words (novice readers) and those who answered correctly by relying on their constructed coherent mental representations of the text (advanced readers) would not be distinguished.

Thus, obtaining evidence that is directly related to the assessed schemas is essential for ensuring the diagnostic validity of assessment tools. A possible approach could be based on directly observing what schemas (if any) learners use immediately when approaching a problem or trying to make sense out of the presented situation. Even though schema-based approaches to the assessment of students and to the design of test items have been suggested before (Marshall 1993, 1995b; Singley and Bennett 2002), the idea of registering rapidly if and how learners use their schemas while they approach a specific problem or situation has a potential value for enhancing cognitive diagnostic assessment (Kalyuga 2006d; Kalyuga and Sweller 2004). A general design methodology and specific implementations of this approach will be described in the following sections.

### **3.2.3 General Design Framework**

The rapid schema-based diagnostic approach is based on observing task-relevant schemas from long-term memory (if any) that are rapidly activated and brought into working memory as learners approach a briefly presented specific task situation. Individuals who are more experienced in the task domain would be better able to recognize presented problem states and retrieve appropriate solution schema steps than less knowledgeable learners. Experts could immediately see a task situation within their higher-level knowledge structures and activate appropriate solution schemas, while novices could only locate some random lower-level components.

The design of a schema-based assessment may follow a general conceptual framework for the design of cognitive assessment containing three basic components: the student model, the task model, and the evidence model (Mislevy et al. 2002). The student model (or model of expertise) describes the cognitive constructs to be assessed, i.e., schemas that guide cognitive processing in a specific task area. The task model defines characteristics and patterns of tasks that would allow obtaining evidence about assessed cognitive knowledge structures. The evidence model defines observable variables, their scoring procedures, and a specific measurement model to be applied to the data.

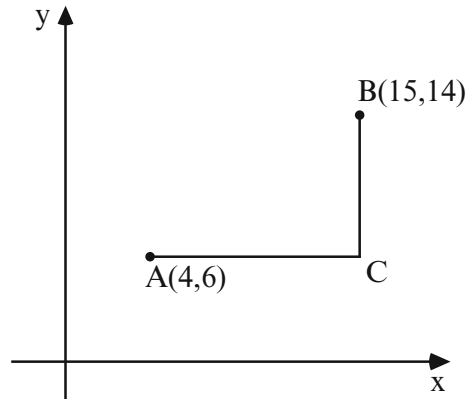
According to this framework, the task-relevant schemas should be described first, followed by a pattern of tasks that would elicit evidence about these schemas, and finally by a scoring procedure for these tasks and a suitable measurement model to make statistical inferences about levels of acquisition of the assessed schemas. The following section describes possible implementations of the above general idea and examples of applying the rapid schema-based assessment to coordinate geometry tasks and arithmetic word problems. These two task areas differ in types of knowledge and levels of knowledge organization.

## **3.3 Rapid Diagnostic Assessment Methods**

The idea of rapid schema-based assessment can be realized either as a first-step method or a rapid verification method. In the first method, learners are presented with a task for a limited time and asked to rapidly indicate their first step toward solution. Different first steps would indicate different levels of expertise. This method was validated in a series of studies using tasks in algebra, coordinate geometry, and arithmetic word problems. Results showed high levels of correlations between performance on the rapid tasks and detailed traditional measures of knowledge (Kalyuga 2006d, 2008; Kalyuga and Sweller 2004), with substantially reduced test times.

The rapid verification method is a version of the first-step procedure designed for the use in computer-based environments. Learners are actually presented with a series of possible steps (some of which are incorrect) at various stages of the solution

**Fig. 3.1** A diagram for the basic task used in the coordinate geometry area (Adapted from Kalyuga and Sweller (2004). Copyright © 2004 by the American Psychological Association, Inc.)



procedure and asked to rapidly verify the correctness of these steps. Knowledge structures of more experienced learners would presumably allow them to verify suggested steps more successfully than novices. This method was validated using sentence comprehension tasks and tasks in kinematics (Kalyuga 2006a, c, 2008). Again, significant correlations were found between performance on the rapid verification tasks and extended traditional measures of expertise, with significantly reduced test times.

Since either of the above two forms of rapid assessment approach could be used with the same student model (model of expertise), task model, and evidence model, the following examples are concentrated on developing and implementing these components of the general design framework using areas of coordinate geometry and arithmetic word problems. According to this framework, a subgoal structure of the tasks and a sequence of corresponding solution steps should be established first. Then, for each step, representative subtasks could be designed and arranged in an appropriate ordered series to be presented to learners, each task for a limited time. The scoring procedure should distinguish student responses corresponding to different levels of expertise in the domain. To assess the level of acquisition of each schema, an appropriate measurement model should be fitted to the data.

### ***3.3.1 Rapid Assessment of Expertise in Coordinate Geometry***

#### **3.3.1.1 Model of Expertise**

A narrow task area selected for demonstrating the method could be described by the basic top-level task (Fig. 3.1) that includes a coordinate plane and two points A and B with given coordinates. Lines AC and BC are parallel to the  $x$ - and  $y$ -axes respectively. The task is to find the lengths of AC and BC. This task effectively

requires finding the distance between projections of two points on a coordinate axis and using the knowledge that opposite sides of a rectangle have equal lengths. The schemas required for solving this task include:

- The schema for determining coordinates of a point as coordinates of projections of the point on  $x$ - and  $y$ -axes
- The schema for establishing equal opposite sides in a rectangle
- The schema for calculating the distance between two points on a coordinate axis (a number line) by subtracting the smaller coordinate value from the larger one (or left hand coordinate from the right hand coordinate, for  $x$ -axis; and lower coordinate from the upper coordinate, for  $y$ -axis)

Different levels of acquisition of these three schemas define the student model in this task area (Kalyuga 2006b). Solving the basic task requires the sequential applications of these schemas to corresponding subtasks. However, a learner who has some schemas at higher levels of acquisition (e.g., automated) could skip some intermediate stages of the solution that would be effectively encapsulated into a higher-level schema. For example, a student who has sufficient prior experience in finding coordinates of a point may find out the coordinates of the points immediately upon presentation of the task without drawing projection lines explicitly. Expert students with extensive experience in this area may immediately (as their first step) write a numerical expression for the length of AC as the difference between  $x$ -coordinates of points B and A.

### 3.3.1.2 Task Model

A pattern of tasks for a rapid assessment of expertise in this task area could have a hierarchical structure with three types of tasks in the pattern: a top-level basic task (requires schemas  $a$ ,  $b$ , and  $c$ ), a task corresponding to the second step in the solution of a basic-level task (requires schemas  $b$  and  $c$ ), and a task corresponding to the final step in the solution of a basic-level task (requires schema  $c$ ). To solve a basic-level task, a learner should acquire schemas necessary for solving each of these three tasks. Lack of any schema would interfere with the entire solution procedure.

Because completing a first step for each task leads directly to one of the following task levels, and each of these levels is represented by another task in the series, the first-step assessment method (or an equivalent rapid verification approach) would diagnose the entire set of schemas in this task area. Accordingly, the tasks should be sequenced according to the number of schemas that are required to solve each of them. For the top-level basic task, no additional details are provided on the diagram. For each of the lower-level tasks, progressively more additional details of partial solutions (e.g., indications of projecting lines and coordinates of the points on axes) are provided on the diagram. For instance, the third task in the series should present most of the details and require only calculation of the differences between the indicated coordinates of two points on each axis.

In task areas like coordinate geometry that use diagrammatic representations as essential components of tasks, each sequential step includes information presented at the previous stages of the solution process. The series of diagnostic tasks in the pattern is effectively a sequence of partially worked-out examples with gradually increasing levels of detail provided to learners. In other domains, it is possible to construct a different task pattern that is based on all possible relevant combinations of basic schemas (see an example for word problems in the next section).

### 3.3.1.3 Evidence Model

In a possible scoring procedure, for each step that requires application of a specific schema, two units are allocated for completing the step and one unit for an intermediate action (an unfinished solution step). If a procedure does not have an intermediate stage, one unit is allocated for completing the step. A zero score is allocated for a wrong answer and for not providing any answer. With a rapid verification method, the same scores are allocated for correct verifications of corresponding solution steps.

For example, for a lower-level task that requires applying only schema *c*, scores 2 and 1 are allocated respectively for providing or verifying a completed final answer ( $AC=11$ ; numbers correspond to Fig. 3.1 for illustrative purposes only, actual diagnostic tasks at different levels should vary in specific numerical parameters) and incomplete final answer ( $AC=15-4$ ).

In contrast, for a top-level task that requires application of all three schemas *a*, *b*, and *c*, scores 5 and 4 are allocated respectively for providing or verifying the above responses at the stages of application of the schema *c* corresponding to the final step and the step that immediately precedes it. A score 3 is allocated for providing or verifying an answer at the stage of application of the second schema *b* (indicating equal sides of a rectangle; there is no intermediate action for this schema). A score 2 is allocated for providing or verifying an answer at the stage of completed application of the first schema *a* (e.g., indicating projections and *x*-coordinates of points A and B). A score 1 is allocated for providing or verifying an intermediate (unfinished) step when applying the first schema (e.g., indicating only a projection line without the coordinate of a point). Thus, an additional score is allocated for each skipped intermediate step in the first-step response (or integrated into an advanced step in the rapid verification procedure)

The application of the first-step method in a paper-based format in a realistic class environment with 20 grade 9 students (Kalyuga and Sweller 2004) indicated a high level of correlation of 0.85 between learners' performance on the rapid test and traditional measures of knowledge of corresponding procedures and concepts, with the test time reduced by a factor of 2.5. The following instructions were provided to students:

In each of the figures, A and B are two points on a coordinate plane. Lines AC and BC are parallel to the coordinate axes. Assume you need to find the lengths of AC and BC.



Some additional details (lines, coordinates) or partial solutions are provided on most figures. For each figure, spend no more than a few seconds to indicate your first step toward solution of the task.

Remember, you do not have to solve the whole task. All you have to do for each figure is to show only your first step toward the solution (e.g., it might be just writing a number or drawing a line on the diagram). If you do not know your answer, proceed to the next page.

Do not spend more than a few seconds for each figure, and do not go back to pages you have already inspected.

### 3.3.2 *Rapid Assessment of Expertise in Solving Arithmetic Word Problems*

#### 3.3.2.1 Model of Expertise

The described model of expertise uses the analysis of schemas in this task area conducted by Marshall (1993, 1995a) that suggested five types of basic schemas. In order to simplify the illustration of the diagnostic method, four of these schemas are used: Change, Group, Vary, and Restate schemas (Kalyuga 2006d).

The Change schema (denoted as *C*-schema for convenience) applies to a situation in which there is a change over time in the value of a variable, for example, *After 5 students had left the class, 12 students remained. How many students were in the class initially?* Students who indicate as their first solution steps (or verify as correct steps) expressions like  $X - 5 = 12$ ,  $5 + 12$ ,  $12 + 5 = 17$ , or  $17$  demonstrate evidence of the Change schema. Different first steps correspond to different levels of the schema acquisition. For example, experienced students may recognize a familiar situation right away and write (or verify) the final answer ( $17$ ) immediately due to their automated schema and do not require much conscious processing in applying this schema.

The Group schema (*G*-schema) relates to situations in which a number of components are combined into a larger unit, for example, *John's homework contains 16 tasks. John completed 11 tasks in the afternoon. In the evening, he did the remaining tasks. How many tasks did John complete in the evening?* Students who write as their first steps or verify expressions  $16 = 11 + X$ ,  $16 - 11$ ,  $16 - 11 = 5$ , or  $5$  demonstrate evidence of the Group schema (on different levels of acquisition).

The Vary schema (*V*-schema) relates to situations in which a systematic relationship exists between two variables: IF the amount of one variable decreases or increases, THEN the amount of the second variable changes in a certain way (*IF-THEN* relationship). The task *A train traveled 120 km in an hour. If the train continued to travel at the same speed, then how far would it travel in 4 h?* requires applying the Vary schema as it could be redescribed as *IF a train traveled 120 km in 1 h, THEN it will travel unknown amount of kilometers in 4 h.* Students who write as their first solution steps or verify statements like  $1 * 4 \rightarrow 120 * 4$ ,  $120 * 4 = 480$ , or  $480$  demonstrate evidence of the Vary schema.

The Restate schema (*R*-schema) applies to situations where there is a known relationship between two variables (ratio-like situations such as *twice as*, *two more than*, etc.) and a restatement of this relationship using different values from those involved in the initial statement, for example, *Water is mixed with cement in the proportion 2:1?. How many units of water are required for 5 units of cement?* Students who write as their first solution steps or verify statements like  $2:1 = X:5$ ,  $5 * 2$ , or  $10$  would demonstrate evidence of the Restate schema.

As previously, the degree of schema acquisition is defined by the level of granularity of solution steps and the number of skipped steps. The levels of acquisition may range from a consciously controlled, slow, and articulated application of all possible fine-grained solution steps (a novice level) to a fluent automated performance with final answers obtained immediately after reading problem statements (an expert level).

The described schema-based model of student expertise is an attempt to impose a schematic structure on a relatively poorly structured task domain using a number of simplifying assumptions. For example, it is assumed that students have sufficient reading comprehension skills that would not introduce an interfering factor. Another assumption is that if a student starts solving a task by drawing a graphical representation, it could be possible to relate unambiguously this diagrammatic representation with a corresponding numerical solution step.

### 3.3.2.2 Task Pattern

Each of the above four basic tasks would require applying only one corresponding schema. There are  $4 \times 4 = 16$  different tasks based on all possible combinations of two schemas. In these combinations, the order of schema applications is important, and repeated applications of the same schema are also allowed.

For example, the task *Paul is thinking of a number. When he adds 6 to the number and then subtracts 9, he would get 15. What is the number John is thinking of?* requires two sequential applications of the *C*-schema (*CC*-task). Applying the first Change schema could result in such responses as  $N - 9 = 15$ ,  $9 + 15$ ,  $9 + 15 = 24$ , or  $24$ . The second Change schema could be used by the students who have completed the first operation, producing the following possible responses:  $N + 6 = 24$ ,  $24 - 6$ ,  $24 - 6 = 18$ , or  $18$ . Some students could also combine two schemas and write  $(15 + 9) - 6$ ,  $15 + 9 - 6$ , etc.

The task *There are 15 boys in a class. The number of girls is 8 more than the number of boys. How many students in the class?* represents an example of the *CG*-task. The Change schema could be applied first with possible responses  $15 + 8$ ,  $15 + 8 = 23$ , or  $23$ . Then, the Group schema could be used with possible responses  $15 + 23$ ,  $15 + 8 + 15$ ,  $(15 + 8) + 15$ , or  $38$ . A *GC* task situation is different from the *CG*-task because it would require applying the Group schema first followed by the application of the Change schema, for example, *Two plates on a table contained respectively 4 and 7 apples. A third plate with apples was added making a total of 18 apples on the table. How many apples were on the third plate?*

Thus, all possible task situations that are based on applying one or two schemas could be represented by a pattern consisting of 20 tasks. Using a similar combinatorial approach, it is also possible to construct three-schema tasks, four-schema tasks, and so on. However, for three-schema tasks, it is unlikely that even highly experienced students would be able to skip first two operational steps and immediately indicate the final third operation or its result as their first step (or immediately verify the final answer). Therefore, a combinatorial pattern of 20 one- and two-schema tasks could be effectively used to collect data on student performance in arithmetic word problem solving.

### 3.3.2.3 Evidence Model

The scoring procedure should reflect different levels of schemas (if any) applied by students while making their first solution step or verifying a suggested step. If the response is based on an immediate next step corresponding to the first schema in the detailed solution sequence for the task, a score 1 should be allocated. If the response is one of the more advanced steps toward the solution (or the final answer), it should be allocated an additional score for each skipped step.

For example, for the above two-schema CG-task, responses at the level of the first schema (*C*-schema),  $15 + 8$  or  $23$ , are scored as 1 or 2 respectively. Responses at the level of the second schema (*G*-schema) such as  $23 + 15$ ,  $15 + 8 + 15$ ,  $(15 + 8) + 15$  are allocated a score 3 (as an intermediate step for the second schema). Responding with (or verifying) the final answer ( $38$ ) would attract a score 4 because three intermediate-level steps were skipped in this case.

In a rapid verification computer-based test, students could be presented the following instructions:

On the following screens, you will see 20 arithmetic problems. You will be allowed a limited time to study each problem.

For each task, several possible (correct or incorrect) solution steps will be presented one at a time. Spend no more than a few seconds to indicate if the provided solution step is correct or incorrect. Click on the “RIGHT” button if you think the step is CORRECT or the “WRONG” button if the step is INCORRECT. If you do not know the answer, click on the “DON’T KNOW” button.

The suggested approach was tested as the first-step technique in a realistic class environment (a paper-based format) with a sample of 55 grade 8 students (Kalyuga 2006d) and compared with a traditional test asking students to write complete solutions to 20 similar (although not identical) problems using a partial credit scoring procedure based on students’ written solutions. The rapid test was 2.8 times faster, with a significant correlation of 0.72 between scores for both tests indicating a sufficient predictive validity of the rapid test.

The traditional classical test theory procedures are usually focused on one-dimensional overall performance indicators. If distinct schemas are defined in the models of student expertise, appropriate multidimensional measurement models could be used to assess each construct separately. In the arithmetic word problems

area, two different multidimensional measurement approaches were applied to the data (Kalyuga 2006b, d). One approach was based on a multidimensional Rasch model (Adams et al. 1997). Another approach was based on Bayesian conditional probabilities estimations using the Markov chain Monte Carlo (MCMC) estimation procedure (Gelman et al. 1995).

In the multidimensional Rasch model, a student's position in the four-dimensional space was defined by a set of four parameters corresponding to four schemas. The ConQuest software for the partial credit model was used to carry out the multidimensional analysis (Wu et al. 1998). Model fit estimates generated by the software indicated acceptable ranges of values for most items. For each student, values of the knowledge variables for each schema dimension and corresponding error variances were determined.

The Bayesian conditional probability model is based on a certain assumption about probabilities  $P(X | S)$  of observing a set of scores  $X$  for 20 tasks if the four-dimensional set  $S$  of a student's knowledge parameters (according to the student model) is known. If some prior hypothetical distribution  $P(S)$  of these variables in the population of interest is defined, it is possible to apply the Bayes theorem to calculate the probability distribution for student parameters conditional on observed test scores,  $P(S | X) \sim P(X | S) P(S)$ . Then, the updated probability distribution could be used as a prior distribution for the next step of updating in the iterative process. For a prior distribution  $P(S)$ , the same categorical distribution for all students and for all four schematic dimensions was defined. The WinBUGS computer program (Bayesian inference Using Gibbs Sampling) was used to estimate posterior distributions conditional on the response data obtained in the experiment (Spiegelhalter et al. 2003). For each student, posterior means and standard deviations for parameters of each schema were estimated.

Although rough and simplified multidimensional methods were used, both models worked reasonably well and produced well-correlated (average correlation of 0.77) estimates of the parameters of students' schemas. Even though these results show that multidimensional measurement models could be used for making statistical inferences about learners' schematic knowledge structures, their application is not always practically plausible in small-scale formative assessments or during training sessions in adaptive instructional systems.

For each learner, a simple data summary using total scores for each schema based on the learner responses to the tasks that involve the corresponding schemas could do equally well. For two-schema items, the score for the first schema could be identical to the entire item score, while the second schema could be scored 1 if the item score is 3, or 2 if the item score is 4. For each of the four schemas, eight tasks contributed to the schema's score (e.g., tasks C, CC, CG, CV, CR, GC, VC, and RC contributed to the C-schema score; the last six items in this set also contributed to other dimensions). The summary scores for each schema dimension correlated significantly (between 0.80 and 0.96) with the parameters for levels of acquisition of corresponding schemas estimated by the two multidimensional measurement models.

### 3.4 Toward Rapid Dynamic Assessment for Learning

The rapid diagnostic methods could be related to dynamic assessment (Bransford and Schwartz 1999; Grigorenko and Sternberg 1998; Sternberg and Grigorenko 2001, 2002). Dynamic assessment is aimed at determining a learner's current stage of development at which he or she can solve a task if a certain level of guidance or help is provided, for example, by showing previous solution steps or hints. For example, if a student fails an item, she could be provided with a hint. If it does not help, another more detailed hint could be presented and the process repeated.

In rapid assessment methods, learners are presented with tasks reflecting various stages of a solution procedure with a gradually changing number of previously completed steps (e.g., see the previously described task model for rapid assessment in coordinate geometry) for making their next step or for rapid verification. Such task sequences effectively represent a form of scaffolding that is used to determine the precise level of learner expertise. This approach also effectively determines the learner zone of proximal development for dynamic selection of learning tasks that are just above the current level of expertise. Integrating the rapid diagnostic assessment approach with dynamic assessment into what could be called rapid dynamic assessment represents an important current direction of research and development in this area.

If learners are presented with incomplete intermediate stages of the task solution and asked to indicate the next step toward solution, they need to recognize both problem states and the solution moves associated with those states. Learners who are more advanced in the domain should be better able to recognize intermediate problem states and retrieve appropriate solution steps than less knowledgeable learners. For example, when training apprentices of manufacturing companies in reading charts used for setting cutting machines (Kalyuga et al. 2000), replacing visual on-screen texts with corresponding auditory explanations was beneficial for novice learners (modality effect). However, when learners became more experienced in using these charts, the best way to present a new type of charts was to display just a diagram without any explanations (an example of the expertise reversal effect Kalyuga et al. (2003)).

An appropriately designed series of rapid dynamic assessment tasks may allow switching instructional formats at the most appropriate time for an individual trainee. Such tasks may include regularly presenting trainees with a series of partially completed procedures in using charts with different degrees of completeness and asking them to indicate their next step toward solution. At the lowest level of completeness, no solution cues or hints are indicated on the chart. At the next level, only some relevant details of the task statement are highlighted. At the following levels, more lines and other solution details are shown. In this way, levels of expertise can be rapidly determined. Less knowledgeable learners then could be presented with comprehensive auditory explanations. In contrast, more experienced trainees, for whom the auditory explanations might be redundant, would learn better from a diagram with limited or no explanations.

Dynamic tests enhance students' learning and, at the same time, provide more accurate measures of current skill levels than traditional static tests. Students learn when they are tested, and they are tested when they learn. Integration of learning and testing into dynamic assessment formats is a current trend in the educational assessment field. For example, Feng et al. (2009) integrated continuous assessment and tutoring in their web-based tutoring system ASSISTment that combined assistance and assessment. The immediate tutoring is provided following each assessment item that students could not solve on their own. In addition to traditional scores based solely on correctness of students' responses to test items, the system collects data on its interactions with students (e.g., time taken to come up with an answer, response accuracy, and speed, time taken to correct an answer if it is wrong, help-seeking behavior as the number of requested hints, and solution attempts on sub-steps) that reflect their effort in solving the test item with instructional assistance in the form of hints, guidance, etc.

If students fail an item, they are provided with a small "tutoring" session where they must answer a few questions that break the problem down into steps. Thus, each ASSISTment task includes an *original question* and a list of *scaffolding questions* to coach students who fail to answer the original one. By analyzing these students' performance on the scaffolding questions, the system provides fine-grained diagnostic information. The system helps students to work through difficult problems by breaking them into sub-steps and meanwhile collecting data on different aspects of student performance (Feng et al. 2009). Thus, instruction is provided to students during the detailed evaluation of their knowledge and skills. As a result, a better evaluation of student abilities and prediction of their future performance is achieved. Since the ASSISTment system automatically provides students with feedback, scaffolding questions, and hints, it provides a form of embedded dynamic assessment.

### 3.5 Conclusion

The general idea of the rapid diagnostic assessment is to determine the level of most advanced domain-specific schemas (if any) a learner is capable of activating immediately on presentation of a test task. This assessment approach essentially evaluates the degree to which the learners' working memory capacity has been expanded due to available schemas in long-term memory. If a more knowledgeable learner is facing a task in a familiar domain, the relevant schemas are rapidly activated allowing the encapsulation of many elements of information (e.g., detailed solution operations and steps) in working memory into a single element (e.g., a higher-level advanced solution step). Different rapid responses would reflect different levels of acquisition of corresponding schemas. Thus, the rapidness of such tests is not only a means of reducing testing time, but it is essential for capturing schemas that learners use in specific situations before they can apply lengthy random search processes and chains of reasoning.

The rapid test tasks could be either used as stand-alone diagnostic probes or presented in a specific sequence. In order to qualify as dynamic assessment tasks, they should be developed as a series with a gradually changing number of completed essential steps or with different levels of instructional support provided in other forms. The diagnostic power of this rapid dynamic assessment may approach that of laboratory-based concurrent verbal reports, however achieved on a considerably shorter time scale.

### ***3.5.1 Future Developments***

#### **3.5.1.1 Establishing Generality of the Tool**

The examples and studies described in this chapter were limited to relatively narrow task areas associated with well-structured problems. In relatively poorly specified domains that involve problems with multiple possible routes to solutions, the rapid verification method could be potentially applied by selecting only a limited number of situations representing different possible paths and levels of solution steps (including both appropriate and unsuitable steps) for rapid verification. The generality and limits of usability of rapid assessment, especially in poorly structured domains, need to be investigated in further research.

In addition to domain-specific schemas, understanding verbally presented problems may also depend on reading comprehension skills and factual knowledge used in specific problem contexts. Therefore, while such tests could be usable with relatively advanced learners (e.g., secondary or high school students) for whom such factors may not influence results, their suitability for less advanced learners (e.g., primary school students) whose responses may depend on a wider range of factors needs to be further investigated.

#### **3.5.1.2 Using Rapid Assessment in Adaptive Learning Environments**

Rapid assessment methods have been applied in adaptive computer-based tutorials for high school students in solving linear algebra equations (Kalyuga and Sweller 2004, 2005) and vector addition motion problems in kinematics (Kalyuga 2006a). The levels of provided instructional guidance in tutorials were based on rapid measures of learner expertise. At the beginning of each session, the initial rapid test was used to select the level of support. For learners with lower levels of expertise, based on the rapid test, additional worked-out examples were provided. For learners with higher levels of expertise, less worked examples and more problem-solving exercises were provided. During the session, rapid tests were used to select the optimal learning pathway. Based on those tests, each learner was either allowed to proceed to the next stage with a lower level of guidance or required to repeat the same stage and then take the rapid test again. At each subsequent stage of the tutoring session,

a lower level of instructional guidance was provided to learners, and a higher level of the rapid diagnostic tasks was used at the end of the stage.

The adaptive tutorials resulted in higher learning outcomes than similar nonadaptive tutorials in which learners either studied all tasks that were included in the corresponding stages of the training session of their yoked participants or were required to study the whole set of tasks available in the tutorial. The described studies provided preliminary evidence for the usability of the rapid assessment methods in adaptive instruction. Similar rapid test-based approaches could be used in other domains (including relatively less structured subject areas) for initial selection of the appropriate formats of learning materials according to levels of users' prior knowledge in the domain, monitoring their progress during learning, and real-time selection of the appropriate learning tasks and instructional formats.

An important direction for further improvements of adaptive learning environments is using rapid dynamic assessment methods (rather than stand-alone rapid tests embedded into the learning sessions) that allow a full and seamless integration of learning and assessment. Rapid dynamic assessment methods could also be used for enhancing assessment oriented toward self-directed learning (Mok 2010) by providing students with real-time evaluation of their current progress in a task domain. To further improve self-directed learning, learner-controlled adaptive environments that provide learner-tailored guidance need to be developed and experimentally tested in future research studies.

## References

- Adams, R., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1–23.
- Baddeley, A. (1997). *Human memory: Theory and practice*. East Sussex: Psychology Press.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61–101). Washington, DC: American Educational Research Association.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Mind, brain, experience, and school*. Washington, DC: National Academy Press.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.
- Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanation: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145–182.
- de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.
- Embretson, S. (1993). Psychometric models for learning and cognitive processes. In N. Frederiksen, R. J. Mисlevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 125–150). Mahwah: Erlbaum.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725–747.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.



- Ericsson, K. A., & Simon, H. A. (1993). Verbal reports as data. *Psychological Review*, *87*, 215–251.
- Feng, M., Heffernan, N., & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, *19*, 243–266.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman & Hall.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, *124*, 75–111.
- Kalyuga, S. (2006a). Assessment of learners' organized knowledge structures in adaptive learning environments. *Applied Cognitive Psychology*, *20*, 333–342.
- Kalyuga, S. (2006b). *Instructing and testing advanced learners: A cognitive load approach*. New York: Nova.
- Kalyuga, S. (2006c). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology*, *26*, 613–627.
- Kalyuga, S. (2006d). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction*, *16*, 1–11.
- Kalyuga, S. (2008). When less is more in cognitive diagnosis: A rapid online method for diagnostic learner task-specific expertise. *Journal of Educational Psychology*, *100*, 603–612.
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, *96*, 558–568.
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, *53*, 83–93.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*, 126–136.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). Expertise reversal effect. *Educational Psychologist*, *38*, 23–31.
- Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). Models of competence in solving physics problems. *Cognitive Science*, *4*, 317–348.
- Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, *13*, 251–283.
- Marshall, S. P. (1993). Assessing schema knowledge. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 155–179). Mahwah: Lawrence Erlbaum Associates.
- Marshall, S. (1995a). *Schemas in problem solving*. New York: Cambridge University Press.
- Marshall, S. (1995b). Some suggestions for alternative assessments. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 431–453). Hillsdale: Lawrence Erlbaum Associates.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, *33*, 379–416.
- Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, *15*, 363–389.
- Mok, M. M. C. (2010). *Self-directed learning oriented assessment: Assessment that informs learning and empowers the learner*. Hong Kong: Pace Publishing.
- 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. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 307–353). Washington, DC: AERA.
- Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Research Council's Committee on the Foundations of Assessment. Washington, DC: National Academy Press.

- Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198.
- Pirolli, P., & Wilson, M. (1998). A theory of the measurement of knowledge content, access, and learning. *Psychological Review*, *105*, 58–82.
- Russell, M., O'Dwyer, L., & Miranda, H. (2009). Diagnosing students' misconceptions in algebra: Results from an experimental pilot study. *Behavior Research Methods*, *41*(2), 414–424.
- Singley, M. K., & Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 361–384). Mahwah: Lawrence Erlbaum Associates.
- Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational measurement* (pp. 263–331). New York: Macmillan.
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS user manual* (Version 1.4) [Computer software]. Cambridge: MRC Biostatistics Unit. Retrieved from <http://www.mrc-bsu.cam.ac.uk/bugs>
- Sternberg, R. J., & Grigorenko, E. L. (2001). All testing is dynamic testing. *Issues in Education*, *7*, 137–170.
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, *32*, 9–31.
- Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.
- Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–487). Hillsdale: Lawrence Erlbaum Associates.
- Van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147–177.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *ConQuest: Generalized item response modeling software [computer program]*. Camberwell: ACER Press.