When Less Is More in Cognitive Diagnosis

Slava Kalyuga (s.kalyuga@unsw.edu.au)
University of New South Wales,
Sydney, 2052 Australia

Abstract

A rapid approach to diagnostic assessment of levels of acquisition of organized domain-specific knowledge structures is described. The approach is based on evaluating an immediate content of long-term working memory and has the potential for developing more rapid and sensitive diagnostic techniques than traditional knowledge tests. To illustrate the approach, a specific rapid diagnostic method in kinematics is described and applied as a means of tailoring instructions to levels of learner expertise in an adaptive computer-based tutor.

Keywords: cognitive diagnosis; rapid diagnostic assessment; adaptive learning; expertise.

Introduction

A range of powerful methods are used in cognitive science for diagnosing individual knowledge structures and other cognitive attributes. Most of these techniques are based on interviews, think-aloud procedures, observations, and analysis of performance records (e.g., Ericsson & Simon, 1993). However, these methods have not been widely used outside laboratory studies because they are very time consuming. Instead, simple traditional testing procedures are used in instructional practice to obtain evidence about learners’ knowledge for diagnostic purposes.

Often, sufficiently fine-grained diagnostic information is required for making decisions in real time, for example, during a single instructional session. For example, in adaptive multimedia and e-learning environments, instructional techniques and materials often need to be adjusted dynamically with alterations in learner expertise (Kalyuga, 2005; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Sweller, in press). It is necessary to have rapid instruments with sufficient diagnostic power for detecting different levels of acquisition of corresponding cognitive constructs. The reported study has been designed to develop an alternative rapid method for diagnosing learners’ knowledge structures in specific domains (involving relatively well-defined problems) based on our current understanding of interactions between working and long-term memory structures.

The way we process information changes as our domain-specific knowledge base develops. Long-term memory (LTM) knowledge structures allow chunking multiple elements of information to effectively reduce or eliminate severe processing limitations of our cognitive system. In addition, we are also able to bypass working memory (WM) limitations by having our knowledge structures in long-term memory highly automated due to extensive practice. Thus, well developed long-term memory knowledge base fundamentally alters characteristics of human cognitive performance (see Sweller, 2003, for a possible explanation of evolutionary advantages of such cognitive architectures).

Studies of expert performance indicate that available domain-specific knowledge enables experts to quickly encode and retain large amounts of information in LTM. Such LTM storage and retrieval operations speed up with practice resulting in experts’ superior task performance and recall for familiar materials. People can be trained to effectively increase their memory capacity to an amazing degree through extensive training in chunking and re-chunking information into meaningful units using their prior knowledge stored in LTM. The skilled memory theory claims that people develop mechanisms that enable them to use their knowledge base to rapidly encode, store, and retrieve information within the area of their expertise and thus circumvent the working memory capacity limitations.

As a result, experts possess an enhanced functional working memory capacity in domains of their expertise (Chase & Ericsson, 1982; Ericsson & Staszewski, 1989).

Ericsson and Kintsch (1995) further developed these ideas into the theory of long-term working memory (LT-WM). In this theory, LTM knowledge structures associated with components of working memory form a LT-WM structure that is capable of holding virtually unlimited amount of information. The proposed mechanism of LT-WM operation involves cue-based retrieval of information from LTM. Skilled performance depends on individual domain-specific knowledge structures relevant to particular tasks, and, consequently, there are individual differences in the operation of LT-WM for a given task (Ericsson & Kintsch, 1995). Thus, working memory capacity limitations do not pose a problem for people who develop expertise in a domain and have their knowledge base well organized in long-term memory.

In diagnostic assessment, we often try to make inferences about learners’ organized knowledge base by analysing results of their problem solving activities using traditional testing procedures. Such data may not always provide reliable and valid diagnostic evidence. For example, observing correct answers to a series of math problems would not tell us how those problems were actually solved: using novice-like random search or trial-and-error method, or using competent application of appropriately organized schematic solution procedures. In the latter case, what level of knowledge was applied: low-level, slow step-by-step procedures or higher-level well-learned and automated...
procedures with rapidly obtained final answers? Even analyses of solution records could not guarantee the valid diagnostic inferences. We do not know what cognitive processes a person was involved before her/his first recorded action or during the breaks between recorded actions.

**Rapid Cognitive Diagnosis: A First-Step Approach**

A more valid approach to making cognitive diagnostic inferences may be based on observing an individual’s immediate use of her/his knowledge during an actual cognitive performance. As mentioned previously, organized knowledge base held in LTM significantly influence the content and characteristics of WM by effectively transforming it into LT-WM. For example, when reading a text in a familiar area, we construct and continuously update a mental representation of the text in WM by retrieving associated components of our knowledge base from LTM. This integrated cognitive construct represents the current content of LT-WM. If someone interrupts our reading with an unrelated conversation, we can resume reading minutes (or even hours) later without loss of comprehension. We often do not need to go again through the sections of the text that had been read prior to the interruption. Due to associations with LTM knowledge base, the content of LT-WM is durable enough and sufficiently resistant to temporary interferences to survive interruptions in reading (see Kintsch, 1998, for details of a theory of reading comprehension).

If a learner is facing a task in a familiar domain, and her or his immediate approach to this task is based on available knowledge structures, these structures will be rapidly activated and brought into the learner’s working memory. A corresponding LTWM structure will be created. Such integrated LTWM knowledge structures define the characteristics of the learner’s working memory during knowledge-based cognitive activities. Observing immediate traces of the content of LTWM while students approach a situation or solve a task could be used to diagnose the level of their expertise in the corresponding domain. These LTWM structures are durable and interference proof to allow sufficient time for a practically usable diagnostic procedure. We are not required to capture the immediate content of WM strictly within a split-second of corresponding cognitive operations. The available time could be sufficient for students to record or otherwise register their responses in a suitable format.

It is practically possible to determine the content of LTWM (or absence of any related content if the person is a novice in the domain) using appropriate procedure and set of cognitive tasks. In a general case, the idea is to determine the highest level of organised knowledge structures (if any) a person is capable of retrieving and applying rapidly to a task or situation she or he encounters. An obvious way of utilizing this idea in practice is to ask students to ‘think aloud’ as they solve a problem, inspect a diagram, or read a text (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Ericsson & Simon, 1993; Magliano & Millis, 2003). Such an assessment procedure, however, could be time consuming and difficult to computerize.

The approach has been realized in an alternative form as the first-step method: learners were presented with a task for a limited time and asked to indicate their first step towards solution (Kalyuga & Sweller, 2004). The first step would involve different cognitive operations for individuals with different levels of expertise in a domain. An expert may immediately provide the final answer; an intermediate level learner can only indicate the very first operation according to a detailed fine-grained solution procedure; and a novice may start some random search process in absence of relevant knowledge of a solution procedure. Different first-step responses would reflect different levels of acquisition of corresponding knowledge structures. When a learner encounters a familiar task, she or he activates an appropriate structure immediately and brings it into WM (create a LTWM structure) to act on the task. Skipping intermediate solution steps would reflect a higher level of proficiency: the learner may have corresponding operations automated or is able to perform them mentally without writing.

In a pilot study involving high school students, the first-step method was used (both in paper- and computer-based formats) to diagnose knowledge of procedures for solving linear algebraic equations, simple coordinate geometry tasks, and arithmetic word problems (Kalyuga & Sweller, 2004; Kalyuga, in press). Experimental results indicated significant correlations (.72 - .92) between performance on these tasks and traditional measures of knowledge that required complete solutions of corresponding tasks. Moreover, test times were reduced by factors of 2.8 - 4.9 in comparison with traditional test times. However, the most important advantage of the rapid diagnostic method was its ability to capture the content of learners’ actual knowledge base when they approach a task. The first-step diagnostic method was not only less time consuming but also more sensitive to underlying knowledge structures than traditional tests. Thus, it has the potential to increase diagnostic power of assessments (approaching that of cognitive laboratory methods) and simultaneously reduce testing time.

In this paper, a different class of tasks from kinematics (vector addition motion problems) is used to study rapid cognitive diagnosis techniques. Consider the following task: *A ship is traveling at 10 m/s. A passenger runs across the deck at 5 m/s in a direction perpendicular to the direction of motion of the ship. Find out the velocity of the passenger relative to the sea.* To solve this problem, the given velocities need to be represented as vectors in a two-dimensional space. Then, a graphical addition operation could be performed on those vectors: \[ V = \sqrt{10^2 + 5^2} \]
When encountered with this task, a student who understands that a vector approach should be applied when directions of movements are not the same or opposite, but who has not practiced graphical addition of vectors, may rapidly start her or his solution by drawing two perpendicular vectors. A student who has more experience with vectors may immediately assign values to the length of each vector. Another student who is familiar with the vector addition procedure may immediately perform the graphical addition. Someone with more experience in adding vectors might be able to write immediately a numerical expression for the Pythagorean Theorem. For a student with substantial experience in solving this class of tasks, such an expression could be the very first operation she or he would write down on paper without even drawing a diagram. The top ‘expert’ in the area could even mentally perform some numeric operations within such expression before writing it down. Learners’ first-step responses to a series of appropriately designed tasks that require knowledge of increasingly larger numbers of solution procedures could provide indicators of the levels of acquisition of corresponding knowledge structures.

If the angles between the velocity vectors are allowed various values, the procedure for calculating the length of the resulting vector may require more advanced knowledge of trigonometry. To limit the study to a relatively simple class of tasks, the range of parameters was intentionally restricted. Angles between vectors could be 0° (the same direction of movements), 90° (perpendicular vectors), 180° (opposite directions of movements), 60°, and 120°. When 60° or 120° angles are used, only equal velocity values for both vectors were allowed. The diagnostic items in this restricted domain could be effectively designed as a sequence of partially worked-out examples with gradually increasing levels of solution details provided to students (a ‘cumulative hierarchy’ pattern). 25 tasks (5 main solution steps X 5 angle values) are included in the suggested task pattern. Each of five groups of tasks is sequenced according to the number of solution steps that have already been completed (from 0 to 4). Within each group, five tasks are sequenced according to the relative direction of movements: same direction, opposite directions, then 90°, 120°, and 60° angles (according to the level of perceived solution difficulty).

Five tasks in the first group provide no information in addition to the textual task statements. The tasks in the second group, in addition to the verbal statements, provide vector graphs indicating only directions of movements. For example, the second task in this group (opposite directions of movements) is A boat is traveling at 7 m/s. A passenger runs at 3 m/s in the direction opposite to the direction of motion of the boat. What is the velocity of the passenger relative to the water?

The tasks in the third group provide vectors graphs with velocity values attached. For example, the third (90° angle) task in the third group is Steady wind of 10 m/s is reported for the area. A bird is flying at 12 m/s in a direction perpendicular to the direction of the wind. What is the velocity of the bird relative to the ground?

The tasks in the fourth group graphically present the vector addition operation. For example, the forth (120° angle) task in this group is A sea wave is traveling at 9 m/s towards the beach. A motorboat moves at 9 m/s in a direction of 120° relative to the direction of the wave. What is the velocity of the boat relative to the ground?

Finally, the tasks in the fifth group provide all necessary graphical information and even indicate a numerical expression for the length of the resulting vector that requires further transformations. For example, the fifth (60° angle) task in the fifth group is A boat is traveling at 6 m/s. A passenger runs across the deck at 6 m/s in a direction of 60° relative to the direction of motion of the boat. What is the velocity of the passenger relative to the water?

By examining a learner’s first steps for each of the tasks, it could be possible to determine the type of approach taken by the student and her or his level of expertise in the domain. The following instructions could be presented to students in a paper-based format: For each of the following 25 tasks, rapidly indicate your first step towards the solution. The first step could be, for example, drawing a diagram, writing a numerical operation, or even providing a final answer (if you can do it immediately).
A possible scoring procedure may allocate a unit score for applying a specific completed procedural solution step. For example, tasks in the fifth group require application of just one step. For each of these tasks, a score 1 would be allocated for providing a final numerical expression or final answer. Tasks in the fourth group require sequential applications of two different procedural steps. For each of these tasks, scores 2 or 1 would be allocated, respectively, for providing a final answer or completing the step that immediately precedes the final operation. Finally, for tasks in the first group that require applications of all five main solution steps, scores 5, 4, 3, 2, or 1 would be allocated, respectively, for providing responses at the stages of completed applications of the corresponding steps.

If a learner omits some intermediate stages, she or he should be allocated an additional unit score for each skipped step. For example, participants who indicate the final answer for a task from the first group (skipping all intermediate steps) would be allocated a score 5 for this task. For the same task, a first-step response that shows a graphical addition of two vectors without any numerical expression for the length of the resulting vector would be allocated a score 3 (two intermediate steps were skipped). Thus, if a learner is experienced enough to indicate the correct final expressions for all 25 tasks, the allocated maximum score would be $5 \times (5 + 4 + 3 + 2 + 1) = 75$.

Rapid Verification Technique

Diagnostic tasks described above and many other tasks in science domains require responses (drawing graphical task representations) that cannot always be specified precisely in advance. In paper-based formats, the first-step method could be applied in all these situations. However, recording and analyzing students’ rapid first-step responses in computer-based environments may be technically challenging. In such situations, an alternative rapid diagnostic approach could be based on students’ rapid verifications of possible suggested solution steps.

With the rapid verification method, learners are presented with a series of suggested possible (correct and incorrect) solution steps reflecting various stages of the solution procedure and are asked to rapidly verify the suggested steps (for example, by pressing corresponding keys on the computer keyboard). For the above vector addition motion tasks, the following instructions could be presented to students on the computer screen: This test contains 25 tasks and takes around 15 minutes. You will be allowed a limited time to study each task. Following each task, different possible solution steps will be presented. For each step you have to immediately 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.

Although response times could be technically limited by allowing students only several seconds to respond, this method might forcefully interrupt some genuine first-step responses. A more suitable alternative approach could be to ‘coach’ students in responding fast before the test. During pre-test exercises with a sample of tasks from a different area, learners could actually see how rapidly they are expected to respond. If a student does not respond within a set time interval, she or he could be asked to respond faster next time. Eventually, when the responses of this student become rapid enough, she or he could be encouraged to respond with this rate during the actual test.

The described pattern of tasks and the rapid verification diagnostic technique were pilot tested with a limited sample of 23 Grade 11 students. Prior to the experiment, during regular classes, students had been taught vector addition methods necessary for solving the tasks included in the test. However, they had not previously encountered the same tasks in the described format. Each of the 25 tasks from the set was presented to students for around 20 seconds. It had been established in pre-trials that this time is sufficient for Grade 11 students to read the statement and inspect the diagram (if included).

Each solution verification window included a diagrammatic and/or numerical representation of a possible (correct or incorrect) solution step and buttons “Right”, “Wrong”, and “Don’t know” for students to click on (see Figure 1 for an example).

![Figure 1: Snapshot of the response window for an item.](image)

Because the number of procedural solution steps required for accomplishing a task decreases from 5 (for the first group of tasks) to 1 (for the fifth group of tasks), different numbers of solution verification questions were allocated for tasks from different groups. For example, for items from the first group, six verification questions (3 correct and 3 incorrect) were provided with gradually increasing levels of diagrammatic or numerical details. Similarly, for items from the second group, five solution step verification questions (e.g., 3 correct and 2 incorrect, or 2 correct and 3 incorrect) were provided. Items from the fifth set required applying one solution step. Therefore, only two numerical options were presented with balanced numbers of correct and incorrect solutions steps across the tasks in the group. The following four statements show examples of four
verification questions that were used for the previously mentioned item from the third group (90° angle):

10 m/s

12 m/s

\[ V = \sqrt{144 + 100} \text{ m/s} \]

10 m/s

\[ V = 22 \text{ m/s} \]

The average total rapid verification test time was 16.5 minutes. An estimate of reliability using Cronbach’s coefficient alpha was .76. For each student, the level of task difficulty at which the solution problems started to appear was identified. The test allowed identifying students with different levels of expertise, in particular those students who had difficulties with more complex vector addition tasks but were able to successfully process tasks that required applying a scalar approach. Thus, the pilot test demonstrated the technique’s diagnostic potential for evaluating different levels of learner expertise in the domain. For further validation purposes, the rapid verification diagnostic method was applied as a means of tailoring instructional procedures to changing levels of learners’ expertise in a simple adaptive computer-based learning environment.

**Method**

The rapid diagnostic technique was used for initial selection of the appropriate instructional materials according to learners’ preliminary knowledge in the domain, as well as for monitoring learners’ progress during instruction and real-time selection of the most appropriate instructional steps. The learner-adapted instructional procedure was compared to an equivalent procedure without real-time adaptation.

**Participants** Sixteen Year 11 (average age of around 17) students from a Sydney school participated in the experiment. The students had not been exposed to the specific materials used in the study prior to the experiment. Students were randomly allocated to the learner-adapted and non-adapted instructional procedures with 8 participants in each group.

**Materials and Procedures** The experiment was conducted in a realistic environment in the school’s computer lab. The instructional packages were designed using Authorware Professional and delivered through desktop PCs. The procedure included an initial rapid diagnostic test (similar to that described above, but with only one task statement and five verification options per angle value), an adaptive training session for the experimental group and a non-adaptive version for the control group, and a final rapid test (similar to the initial test with re-worded tasks).

The training session was based on a series of faded worked examples or completion tasks (Renkl & Atkinson, 2003; Van Merriënboer, 1990) each followed by a problem-solving practice. According to this approach, novices learn most effectively when instructed using fully worked out examples. As levels of learners’ knowledge in the domain increases, parts of worked examples could be gradually omitted thus increasing a relative share of problem solving practice in instruction.

In the learner-adapted group, learners were allocated to appropriate stages of the instructional procedure corresponding to the performance break-down points that were determined by the outcomes of the initial rapid diagnostic test. Appropriate fully and partially worked-out examples were presented, each followed by a problem solving exercise. Time spent studying worked examples was user controlled and time for solving a problem was limited to 1 minute. If all attempts within this time limit were unsuccessful, learners were presented with a fully worked out solution of the problem.

To monitor individual learners’ progress, single rapid diagnostic tasks similar to the tasks at the corresponding levels of the initial diagnostic test were used. Depending on the outcome of these rapid knowledge probes, the learner was allowed to proceed to the next stage of the training session or was required to repeat the same stage and then take the rapid stage test again. Each stage of the training session for a specific sub-class of tasks was similar to the previous stage except for a lower level of instructional guidance provided to learners (in faded examples, increasingly more explanations of initial procedural steps were eliminated) and a higher level of the rapid test task at the end of the stage. How long each learner stayed at each stage depended on her or his performance on rapid diagnostic tasks during the session.

In contrast, in the non-learner-adapted group, all learners went through all the stages of the training session regardless of their performance on the initial rapid diagnostic test. Each learner had to study all worked examples, perform all problem exercises, and undertake all rapid diagnostic tests.
that were included in the training session (however, the outcomes of the rapid diagnostic tests during the session were not used for selecting the subsequent instructional materials).

**Results**

The independent variable was the format of the training session (learner-adapted or non-adapted). The dependent variables under analysis were differences between the sum of the test scores for the final rapid test and sum of the test scores for the initial rapid test (as indicators of learners’ knowledge gains due to the training session), and training session time.

The learner-adapted group indicated better knowledge gains ($M = 19.2, SD = 12.6$) than the non-adapted group ($M = 12.7, SD = 11.7$). The effect size, conservatively estimated using the higher standard deviation value, was .52 indicating a medium size effect. There was no statistically significant difference between the groups in knowledge gains (possibly due to a small sample size). However, there was a significant difference for training session time, $M = 1960, SD = 628$, for the learner-adapted format and $M = 2797, SD = 1154$, for the non-adapted format), $t(14) = -2.27$, effect size = .73.

**Discussion**

The higher knowledge gains for the learner-adapted instructional format in comparison with the non-adapted format of training, together with significantly reduced training time provide strong evidence that the suggested rapid technique for diagnosing learner levels of expertise can be successfully used to individualize instructional procedures. The rapid verification approach potentially allows designing computerized rapid on-line diagnostic assessments for practically any task domain and any type of knowledge (and using these testing techniques in on-line instructional systems to tailor instructional procedures and materials to changing levels of learner expertise).

For example, in relatively less well-defined domains that involve multiple-step problems, students might be able to take different routes to problem solutions. If too many routes are potentially available, the first-step method may not be feasible. However, with the rapid verification technique, it is possible to select only a limited number of steps representing different levels of problem solution procedures. Then, the levels of expertise could be assessed by asking participants to verify rapidly each of the sequentially presented selected solution steps.

The described preliminary studies indicate that, in comparison with traditional tests, the rapid approach has the potential to provide more diagnostic information in less time. To further validate the rapid diagnostic approach (both the first-step and rapid verification methods) it is necessary to test its generality and limits of usability by applying the described methods in other areas and comparing them to more in-depth traditional cognitive diagnostic approaches (e.g., think-aloud protocols).

**References**


