Rapid Dynamic Assessment of Expertise to Improve the Efficiency of Adaptive E-learning

□ Slava Kalyuga
John Sweller

In this article we suggest a method of evaluating learner expertise based on assessment of the content of working memory and the extent to which cognitive load has been reduced by knowledge retrieved from long-term memory. The method was tested in an experiment with an elementary algebra tutor using a yoked control design. In the learner-adapted experimental group, instruction was dynamically tailored to changing levels of expertise using rapid tests of knowledge combined with measures of cognitive load. In the nonadapted control group, each learner was exposed to exactly the same instructional procedures as those experienced by the learner's yoked participant. The experimental group demonstrated higher knowledge and cognitive efficiency gains than the control group.

☐ Processing limitations of human working memory are known to be a major factor influencing the effectiveness of instructional presentations. A limited working memory capacity could easily be overloaded if more than a few chunks of information are processed at the same time (e.g., Baddeley, 1986; Miller, 1956). Organdomain-specific, long-term memory knowledge structures (or schemas) allow people to overcome the limitations of working memory by "chunking" many elements of information into a single, higher-level element (see Chi, Glaser, & Rees, 1982; Larkin, McDermott, Simon, & Simon, 1980). By treating many elements of information as a single element in working memory, long-term memory schematic knowledge structures may reduce working memory load. Another way to reduce working memory processing limitations is to practice the skills provided by schemas until they can operate under automatic rather than controlled processing (Kotovsky, Hayes, & Simon, 1985; Shiffrin & Schneider, 1977).

From a cognitive load perspective, the major role of learning in cognitive functioning is acquisition and automation of schematic knowledge structures in long-term memory. As our previous research on the expertise reversal effect indicated (see Kalyuga, Ayres, Chandler, & Sweller, 2003 for an overview), the effectiveness of learning from different instructional formats and procedures may significantly alter with the development of learner expertise in a domain, because of cognitive load factors. As learners acquire more knowledge in a domain, their prevailing cognitive activities change. Construction of new schemas is the novice's dominant cognitive activity that requires appropriate instructional support. In contrast, experts tend to

retrieve and apply available long-term memory knowledge structures to handle situations and tasks within their area of expertise. The coordination of cognitive activities and control of the flow of information at different stages are controlled by a central executive, according to some theorists. For example, the concept of a central executive developed by Baddeley (1986) explains working memory operations in lower-level cognitive tasks. In realistic educational settings, as well as in most everyday situations, people are usually dealing with complex cognitive tasks that require extensive use of available long-term memory knowledge.

In general, a central executive should be able to determine what information learners attend to and what cognitive activities they engage in at any specific stage of cognition. Sweller (2003), in a departure from current theories, suggested that such decisions must be based entirely on a combination of knowledge and random factors. If there are no other factors available, a fixed central executive is excluded. Within the present framework, well-developed schemas held in long-term memory are a major source of knowledge able to carry out an executive function during high-level cognitive processes. That is, when dealing with high-level cognitive processes, we consider a central executive not as a permanent processor in working memory, but as an entity constructed for every specific task at hand by retrieving appropriate schemas from long-term memory. Such a functional approach regards a central executive as a set of schemas temporarily combined to perform a specific function of managing incoming information streams for a specific task. Because the functional central executive operates within working memory, it obviously consumes working resources, unless all relevant schemas are fully automated.

When faced with a novel situation, people use general search strategies in an attempt to impose some order on elements of incoming information. Alternatively, instructional explanations can also perform an executive function when dealing with unlearned material. Providing a partial substitute for the missing schemabased cognitive central executive at the initial stage of learning might be a primary function of

instruction. Direct instructional guidance (e.g., worked examples) is able to provide a strong substitute for a cognitive central executive: It tells the learner exactly how to manipulate the information to solve a task. In contrast, problem solving or exploratory learning provide the least effective executive function at the initial stages of learning (Kalyuga, Chandler, Tuovinen, & Sweller, 2001).

The relative weight of schemas and instructional explanations in a learner's central executive for a task depends on the level of learner expertise. For novices in a domain who are facing a new task, instruction provides the only available guidance. For experts in the domain who are facing the same task, the task will be very familiar and all necessary schemas are likely to be available in long-term memory. At intermediate levels of knowledge, these two components complement each other, ideally without gaps or overlap. A well-balanced central executive should include all necessary schemas for dealing with previously learned, lower-level elements of incoming information, and full instructional guidance for dealing with all unlearned, higher-level units.

One obvious condition under which a central executive is unbalanced occurs if there are elements of information for which no guidance is provided by either component of the central executive. When dealing with these elements of information, learners have to resort to problemsolving search strategies that are cognitively inefficient and can cause high working-memory loads (Sweller, 1988). The central executive might also be unbalanced if there were an overlap between its schema-based and instructionbased components, because both components are available for dealing with some units of information. In this case, the learner is likely to attempt to relate the overlapping components of the central executive. This process of cross-referencing the related components of the central executive will require additional working-memory resources. As a result, less cognitive capacity will be left for schema application and automation. Such an unbalanced central executive is a result of nonoptimal cognitive load management during instruction and represents a basic cause of the expertise reversal effect.

In order to optimize cognitive load, instructional designs should be tailored to changing levels of learner expertise in a specific domain. User-tailored instructional procedures require methods of assessment of learner expertise that could be used in real time during instruction. Because long-term memory schemas in their guiding, executive role define the features and processes of working memory during knowledge-based performances, evaluating the executive components of working memory during complex cognitive activities may provide indicators of levels of expertise. According to the theory of long-term working memory (Ericsson & Kintsch, 1995), experts' schemas associated with components of working memory create longterm working memory structures with characteristics that are different from those of working memory in knowledge-lean situations. Longterm working memory has virtually unlimited capacity and prolonged duration. The theory of long-term working memory describes a possible mechanism of executive functioning of schemas held in long-term memory. From this point of view, assessment of levels of expertise in a domain should be based on tests of the content of a student's long-term working memory (if any) during complex cognitive activities. How can researchers capture that content?

Generally, educators can present students with a task in a domain to try to find the highest level of schematic knowledge that students apply to the task. High-knowledge individuals may be able to see higher-level structures in the material and analyze it using a conceptually driven approach. Novices at best may see only some randomly combined lower-level components and apply a data-driven analysis. The availability of schema-based knowledge structures in long-term memory is the major factor determining such expert-novice differences.

In the case of procedural knowledge, researchers need to establish the schematic knowledge that guides the students' problem solving as they approach a task. We argue that this information can be obtained by asking students to rapidly indicate their first step toward solution of a task. Experts should be able to rapidly retrieve appropriate solution schemas. For example, when encountering the algebraic equa-

tion $(2x - 3) \div 2 = 3$, some students (depending on their level of expertise) might first multiply both sides of the equation by 2 to obtain 2x - 3 =6; others could effortlessly and rapidly do that operation mentally while adding 3 to both sides of the equation, generating 2x = 9 as their first recorded step; high-level experts could also mentally divide both sides by 2 and write the final answer as their first step: $x = 9 \div 2$. Such step skipping is an important characteristic of higher levels of expertise as a consequence of well learned or automated solution procedures (Blessing & Anderson, 1996; Sweller, Mawer, & Ward, 1983). On the other hand, novices may at best apply some general problem-solving methods, such as means-ends analysis or a trial and error technique, which could also be evident from their first solution step.

Our preliminary studies (Kalyuga & Sweller, 2004) using algebra and coordinate geometry materials for Year 9 and 10 students indicated highly significant correlations (up to .92) between performance on such tests and traditional measures of knowledge that required complete solutions of corresponding tasks. The purpose of the present study was to find out if the suggested testing approach could be implemented in a computer-based learning environment for dynamically adapting instruction to changing levels of learner expertise in a domain.

The method was tested in an experiment with an algebra tutor using a yoked control design. The computer-based tutor in elementary algebra equations was designed as a series of worked examples, completion assignments, and conventional problems combined using the completion strategy (van Merriënboer, 1990). This strategy is based on a sequence of instructional procedures from fully worked out examples with complete task solutions conventional problems. Completion assignments contain a problem description, incomplete solution, and tasks to complete. Renkl and Atkinson (2003) demonstrated that as levels of learner knowledge in a domain increase, parts of worked examples should be progressively replaced with problem-solving steps (faded worked examples technique). In the learneradapted experimental group of the current

experiment, the allocation of learners to appropriate completion assignments or stages of the faded worked examples was based on the outcomes of an initial, rapid, diagnostic test; the learners' progress through the training session was monitored, and instruction was tailored according to changing levels of expertise, again using rapid tests of expertise. In the nonadapted control group, each learner was exposed to the same instructional procedures as those experienced by the learner's yoked participant but was not tested during the training session.

To produce a better indicator of the learners' expertise in a domain, the rapid measures of schematic knowledge were combined with measures of cognitive load based on subjective ratings of mental effort. Higher levels of expertise are associated with more effortless performance of corresponding tasks because of a partial or full automation of well-learned procedures. Cognitive efficiency as a combined measure of the rapid test performance and mental effort was used in the completion strategy for the initial selection of the appropriate assignments, as well as for monitoring learners' progress during instruction and real-time tailoring of instruction to changing levels of expertise. We hypothesized that by the end of the session, the learneradapted experimental group would demonstrate better instructional outcomes than the nonadapted group.

METHOD

Participants

Thirty Grade 10 students from a Sydney school participated in the experiment. At the time of the experiment, students had been taught algebraic operations and equations necessary for solving tasks included in the test. The participants were randomly assigned to 15 pairs; for each pair, one student was randomly assigned to the experimental group and the other was assigned to the yoked control group.

Materials and Procedure

Instructional packages were designed using Authorware Professional software and delivered through four Power Macintosh desktop computers. All participants were trained and tested individually in one session. Initially, each participant was presented with exercises in typing simple algebraic expressions, including all operations and symbols that learners could encounter in subsequent sessions. The experimental procedure included an initial rapid diagnostic test, an adaptive training session for the experimental group with yoked controls in the control group, and a final rapid diagnostic test.

Initial rapid diagnostic test. The initial test was designed to assess the initial level of learner expertise in the domain. In the task statement preceding the test, students were asked, for each equation they would see on the following pages, to type a single one-line step that they would normally make first when solving the equation on paper. A specific example using the equation 2(3x - 1) = 1 was provided. It was explained that when asked to solve this equation, some people would first write 2 * 3x - 2 * 1 = 1, others could start from 6x - 2 = 1 or 6x = 3, and some might even write the final answer ($x = \frac{1}{2}$) as their first step. If they did not know how to solve an equation, students were instructed to click DON'T KNOW. No more than one minute was allowed to type each answer.

Three equations were presented on the following three pages. For the first equation, -3x =7, a score of 2 was allocated for typing the final answer, or 1 for typing an intermediate solution step, for example, $-3x \div 3 = 7 \div 3$. A nil score was allocated for a wrong answer, no response, or for pressing the DON'T KNOW button. For the second equation, 4x + 3 = 2, scores of 4 or 3 were allocated respectively for typing the final answer or the step immediately preceding it. A score of 2 was given for an answer at the level of the first equation (e.g., 4x = -1 or 4x = 2 - 3), and a score of 1 for an intermediate step preceding it, for example, 4x + 3 - 3 = 2 - 3. Similarly, for the third equation, $(2x + 1) \div 3 = 2$, scores ranging from 6 to 0 were allocated (e.g., 6 for the final answer, $x = 5 \div 2$, 5 for typing $2x \div 2 = 5 \div 2$, 4 for typing 2x = 5 or 2x = 6 - 1, etc.).

After learners typed their first solution step, the following instruction appeared on the screen: "Indicate how difficult this task was by clicking on an appropriate answer." The participants had to select one of the nine options (Extremely easy, Very easy, Moderately easy, Slightly easy, Neither easy nor difficult, Slightly difficult, Moderately difficult, Very difficult, and Extremely difficult). Thus, a mental effort rating ranging from 1 (extremely easy) to 9 (extremely difficult) was collected for each task.

The mental effort rating (R) of each task was combined with the performance measure on the same task (P) to provide an indicator of cognitive efficiency (E), where E was defined by E = P÷ R. This definition was different from that first suggested by Paas and van Merriënboer (1993) and used in our previous studies (e.g., Kalyuga et al., 2001). In those studies, efficiency indicators were calculated after experiments had been completed. The efficiency was effectively defined as the difference between z-scores for performance and mental effort ratings, and the experimental data means and standard deviations were used to calculate these z-scores. In the current study, we needed efficiency indicators in real time during the experiment, and so could not calculate z-scores in a similar way. Nevertheless, the efficiency indicator, defined as $P \div R$, has similar general features to those of the previously used construct in that efficiency is higher if similar levels of performance are reached with less effort or, alternatively, higher levels of performance are reached with the same mental effort invested. This definition is consistent with the common sense understanding of efficiency as an effect relative to the resources spent to obtain the result.

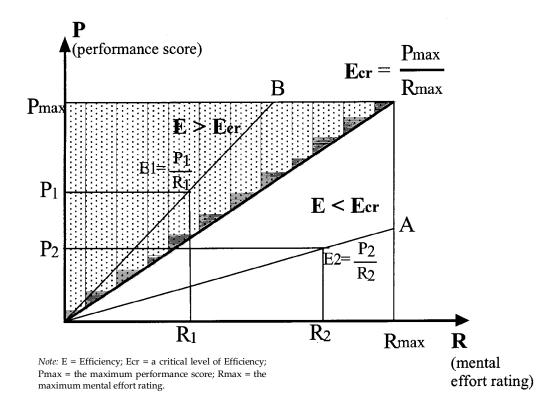
For each task level, a critical level of cognitive efficiency was defined as $Ecr = Pmax \div Rmax$, or $Ecr = Pmax \div 9$, where Pmax is the maximum performance score for the given task level. Cognitive performance of a learner in a given task was considered efficient (and the learner was regarded as competent), if E > Ecr (see Figure 1). If $E \le Ecr$, cognitive performance was regarded as relatively inefficient and the learner was regarded as less competent. Although the values of the critical efficiency levels are selected arbitrarily, they could be refined or adjusted experimentally. The rationale for such levels was based on the general assumption that if someone invests maximum mental effort in a task but

does not display the maximum level of the task performance, his or her cognitive performance should not be regarded as efficient (e.g., point A in Figure 1). On the other hand, if someone performs at the maximum level with less than a maximum mental effort, his or her cognitive performance should be regarded as efficient (e.g., point B). All other (nonextreme) cases (e.g., points R1,P1 and R2,P2) should be judged relative to the critical level. Similar to Paas and van Merriënboer's (1993) approach, the proposed definition is based on a simplifying assumption of a linear relation between performance and mental effort (which does not hold well when performance approaches asymptotic levels). The critical levels of efficiency for the first, second, and third tasks in the rapid test are correspondingly ²/₉, ⁴/₉, and ⁶/₉ (or ²/₃). These critical levels were used for allocating learners in the experimental group to appropriate stages in the training session.

Training session. The training session was designed as a series of worked examples, completion assignments (faded worked examples), and conventional problems, according to the completion strategy. In the learner-adapted format, the allocation of learners to appropriate stages of instruction was based on the outcomes of the initial rapid diagnostic test (Figure 2). Learners who obtained an efficiency level of 3/9 or less on the first task of the initial diagnostic test started training from the first stage. Two fully 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 3 min. If all attempts within the 3-min limit were unsuccessful, learners were presented with a fully worked out solution.

A rapid diagnostic procedure similar to that in the initial diagnostic test was used for monitoring learner progress during the training session (with the equation -4x = 3). In order to be able to get to the next stage of training, a learner efficiency indicator had to be more than $\frac{2}{9}$. If the efficiency indicator was equal to or less than $\frac{2}{9}$ but more than $\frac{1}{9}$, the learner had to undertake some additional training in the form of a set of four shortened worked examples that demon-





strated the major steps, without detailed explanations of intermediate procedures, followed by another attempt on the rapid diagnostic task. If the efficiency was equal to or less than ½, the learner had to go through more training, including two fully worked out examples followed by the set of four shortened worked examples, and then attempt the diagnostic task again.

If a learner efficiency measure on the first task of the initial diagnostic test was more than 2%, but on the second task was 4% or less, the learner started the training session from the second stage, which contained two completion assignments (faded worked examples), each followed by a corresponding problem exercise. In completion assignments, the explanation of the last procedural step was omitted, and learners were asked to complete the solution themselves and type in their final answer. If a learner could

not solve the remaining equation (of the type, 2x = 5) in one minute, the correct solution was provided. In problem-solving exercises, if learners' attempts within the 3-min limit were unsuccessful, they were presented with a fully worked out solution. At the end of this training stage, a rapid diagnostic test similar to the second question of the initial diagnostic test (3x + 7 = 3) was used. The procedure followed was identical to the procedure for the first stage (Figure 2).

If a learner efficiency indicator on the first task of the initial diagnostic test was more than $\frac{2}{9}$, and on the second task was more than $\frac{4}{9}$, but on the third task was $\frac{2}{3}$ or less, the learner started the training session from the third stage. This stage was similar to the previous one, except that a lower level of instructional guidance was provided to learners (in completion assignments, explanations of the two final pro-

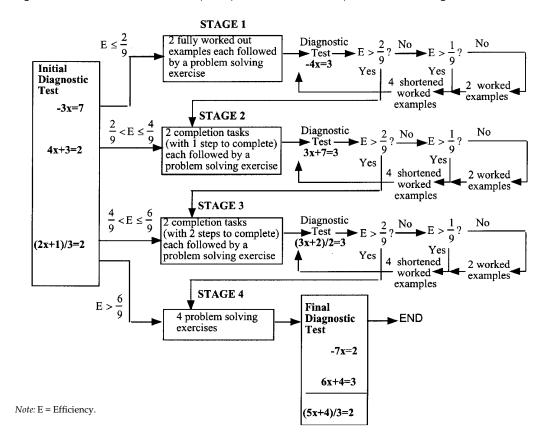


Figure 2 | Flow chart of the adaptive procedure for the experimental training session.

cedural steps were eliminated). At the end of this stage, a rapid diagnostic test similar to the third question of the initial test was used. Finally, if a learner's efficiency indicators on all three tasks of the initial diagnostic test were above corresponding critical levels (indicating an expert performance level), the learner started training from the final fourth stage, which contained only four problem-solving exercises with corresponding feedback.

Thus, in the learner-adapted format, learners who indicated a low cognitive efficiency level in the initial diagnostic test went through all four stages of the training session. How long they stayed at each stage depended on their performance on diagnostic tests during the session. At the other extreme, learners who indicated the highest cognitive efficiency level in the initial

diagnostic test were expected to start the training session from the fourth stage, which included only problem-solving exercises. In contrast, in the nonadapted format group, each learner started the training session from the same stage as the paired learner in the learneradapted format group, and performance during the training was not monitored. The learners in both groups went through the same stages of the training session, thus equalizing experimental conditions. The only difference was that in the learner-adapted format group, the selection and sequence of training episodes was tailored to the specific current level of each learner's expertise, whereas in the nonadapted format group it was not (it was tailored to the level of a different individual, the correspondent yoked participant).

Final rapid diagnostic test. After learners completed the training session, they were presented with the final rapid diagnostic test, which was similar to the initial diagnostic test (Figure 2).

RESULTS

The independent variable was the format of the training session (learner adapted or non-adapted). The dependent variables under analysis were:

- Initial test scores (the sum of the three test scores for the initial test).
- Final test scores (the sum of the three test scores for the final test).
- Differences between the final and initial test scores as indicators of the gains in learner knowledge due to the instructional format.
- Differences between average (over the three test items) cognitive efficiency scores for the final rapid test and average cognitive efficiency scores for the initial rapid test providing indicators of gains in learners cognitive efficiency due to the format of the training session.
- Instruction time.

Means and standard deviations are displayed in Table 1.

There was a significant difference between groups for average efficiency gains, t(14) = 1.89, p < 0.05 (one-tailed t tests for dependent groups were conducted because all hypotheses were directional), Cohen's d effect size 0.69. The learner-adapted instructional format resulted in significantly higher average efficiency gains than the nonadapted (yoked-control) format.¹ There were marginally significant differences for test score gains, t(14) = 1.51, p < 0.1, Cohen's d effect size 0.55. The means indicated a higher test score gain for the learner-adapted group (a medium-to-large effect size).²

Table 1 Mean test scores, knowledge and efficiency gains, and instruction time by experimental conditions.

	Experimental conditions	
	Learner-	Non-
	adapted	adapted
Variable	format	format
Initial total test score		
M	3.00	4.00
SD	1.81	2.93
Final total test score		
M	5.27	4.67
SD	3.26	3.35
Knowledge gains		
M	2.27	0.67
SD	2.28	3.52
Efficiency gains		
M	0.63	0.03
SD	1.03	0.66
Instruction time		
M	708.00	725.00
SD	352.00	268.00

Although training stages and times in both formats were equalized by the use of yoked pairs, the learner-adapted format included intermediate diagnostic test items, each followed by a separate rating scale, that were not included in the nonadapted format. Many learners in the learner-adapted condition had to go through the same diagnostic items and rating scales several times before they attained a sufficient level of cognitive efficiency to proceed to the next training stage. This characteristic of the learneradapted format resulted in an extension of the total training session time for this group. When the time related to diagnostic test items and rating scales (noninstruction time) was deducted from the total training session time (Table 1 shows instruction times without the time required by the diagnostic tests and rating scales), the mean instruction times for the two groups were identical due to the equalizing function of the yoked pairs procedure.

Student actions during learning were recorded electronically and indicated that most participants in the learner-adapted group proceeded through all four stages of the training session (because they were novices); 2 partici-

¹ When the three items were analyzed separately, there were no statistically significant differences in efficiency gains for the first two items, although the gains were always higher for the learner-adapted group; there was a significant difference for the third item, t(14) = 2.11, p < .025.

^{2.} When the three items were analyzed separately, there were marginally significant differences in item score gains for the first (t(14) = 1.47, p < .1) and third (t(14) = 1.48, p < .1) items, and no difference for the second item.

pants started their training from the second stage. The individual progression of learning in the learner-adapted group varied considerably. For example, 11 participants proceeded through the Stage 1 diagnostic test more than once, and all of them studied both full and shortened worked examples at least once; 8 participants went through this test more than twice, 5 participants repeated it more than three times, and 3 participants proceeded through this test more than four times. All participants went through the Stage 2 diagnostic test only once, and 4 participants proceeded through the Stage 3 test more than once (2 of them studied only shortened worked examples). This set of sequences indicates that the first operation (dividing both parts of an equation by the same number) was the most difficult one for many students, and the second operation (adding the same number to both sides of an equation) was the easiest one. Only 3 students completed all problem exercises and completion assignments without errors and repeated attempts. Three participants reattempted some solutions at each of four stages, 4 participants made repeated attempts on at least three stages, 3 students reattempted solutions at two stages, and 2 participants reattempted solutions only at one stage. Together with higher efficiency gains and somewhat higher knowledge gains for the learner-adapted group, these observations show that the adaptive procedure was successful in individualizing instructional sequences during the training session.

DISCUSSION

Optimizing cognitive load and building adaptive e-learning environments require new rapid methods of assessment of student levels of knowledge. In this article, a rapid method of measuring learner levels of expertise in a specific area was used together with online measures of subjective mental effort for the dynamic selection of learning tasks. Students were asked to indicate their first step toward solution of each equation and rate their mental effort while performing the task. More expert learners were assumed to use their schematic knowledge base to retrieve appropriate solution schemas with

minimal effort. Experimental results indicate that the method could be used to build efficient learner-adapted instructional procedures in electronic learning environments based on monitoring learner ongoing cognitive performance. The suggested approach may also contribute to implementing principles of deliberate practice in online learning environments for advanced learners (see van Gog, Ericsson, Rikers, & Paas, this issue).

The yoked control design allowed us to isolate the effects of adaptive training separately from other possible factors influencing instructional efficiency, such as the number of studied examples and exercises, or instruction time. However, the yoked treatment might result in inappropriate training schedules for some participants, for instance, when a low-knowledge learner happens to be a yoked control for a highknowledge learner. Stronger control and treatment comparisons may be needed in future studies; for example, a comparison of a learneradapted condition with an instructional procedure containing all possible examples and exercises giving a maximum training treatment. Also, the adaptive decision rules employed in this study were selected in a rather arbitrary fashion based on commonsense considerations. Future empirical studies and comparisons of different decision rules could provide better algorithms and increase the instructional effects of adaptive training.

It should be noted that the learners in the adapted format group sometimes repeated the same diagnostic items and additional training segments (containing sets of two full worked examples and/or four shortened worked examples) several times, until their efficiency indicator allowed them to continue to the next training phase. Their yoked controls received the corresponding sets of fully worked examples and shortened worked examples once only, although the time made available to the controls was equal to the total time available to the adapted format group. Thus, for some pairs, the adapted format group received the same material for relatively short periods on multiple occasions, while the yoked controls received that material once only, but for a relatively longer period that equaled the addition of the multiple

periods required by the adapted format group. It is possible that the repetition, and not only the adaptability, caused the higher efficiency and knowledge gains for the experimental group. In future research, this factor could be experimentally controlled for.

The constructed training program was dynamic, whole-task based (students learned to solve equations, not separate operations), and individualized. It incorporated cognitive load as an essential factor of the learning task selection procedure. An important advantage of the suggested approach for building learner-adapted instructional e-learning environments is its relative simplicity. Using rapid diagnostic tests for real-time online monitoring of student performance does not require all the complex computational machinery of continuous tracing of student models used in most intelligent tutoring systems. We applied a dynamic-problem-selection approach to the completion strategy similar to that implemented in CASCO (completion assignment constructor; van Merriënboer, Krammer, & Maaswinkel, 1994), but without using a complicated fuzzy-logic-based instructional model and student profile.

The dynamic-problem-selection procedure realized in this article is also similar to those implemented by Camp, Paas, Rikers, and van Merriënboer (2001) and Salden, Paas, Broers, and van Merriënboer (2004) in air traffic control training. However, we used an entirely different performance assessment method (based on rapid cognitive tests), an alternative definition of instructional efficiency, and a different task selection algorithm that allowed a smoother transition between stages of training. Despite these differences in approach, learner-adapted formats proved to be more effective than non-adapted formats in all these studies.

This study was designed to test the usability of the learner-adapted procedure in e-learning environments. The suggested efficiency-based approach requires further investigation and fine-tuning by, for example, using varying criteria of attaining required levels of knowledge in a domain, and designing less intrusive nonverbal formats of subjective rating scales more suitable for electronic learning environments. A motivational perspective on the relation

between mental effort and performance could also be taken into account (see Paas, Tuovinen, & van Merriënboer, this issue). To determine the limits of applicability of the approach in adaptive e-learning environments, we need to test it in other domains, especially in less structured areas such as monitoring language comprehension in online reading tutors.

Slava Kalyuga [s.kalyuga@unsw.edu.au] is with Educational Assessment Australia at the University of New South Wales, Sydney, Australia.

John Sweller is with the School of Education at the University of New South Wales.

This article was prepared during the appointment of the first author as an adjunct lecturer at the School of Education, University of New South Wales.

Correspondence concerning this article should be addressed to Slava Kalyuga, School of Education, University of New South Wales, NSW 2052 Australia.

REFERENCES

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 576–598.

Camp, G., Paas, F., Rikers, R., & van Merriënboer, J. J. G. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort, and mental efficiency. Computers in Human Behavior, 17, 575–595.

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). Expertise reversal effect. *Educational Psychologist*, 38, 23–31.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychol*ogy, 93, 579–588.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96, 558–568.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problem hard? Evidence from Tower of Hanoi. Cognitive Psychology, 17, 248–294.

Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980). Models of competence in solving physics problems. Cognitive Science, 4, 317–348.

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.
- Paas, F., Tuovinen, J. E., & van Merriënboer, J. J. G. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learners' involvement in instructional conditions. [This special issue]. Educational Technology Research and Development, 53(3), 25–33.
- Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. Human Factors, 35, 737–743.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist*, 38, 15–22.
- Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science*, 32, 153–172.
- Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257–

- 285.
- Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), The psychology of learning and motivation, Vol. 43 (pp. 215–266). San Diego, CA: Academic Press.
- Sweller, J., Mawer, R., & Ward, M. (1983). Development of expertise in mathematical problem solving. Journal of Experimental Psychology: General, 12, 639–661.
- van Gog, T., Ericsson, K. A., Rikers, R. M. J. P., & Paas, F. (2005). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. [This special issue]. *Educational Technology Research and Development*, 53(3), 73–81.
- van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research*, *6*, 265–287.
- van Merriënboer, J. J. G., Krammer, H. P. M., & Maaswinkel, R. M. (1994). Automating the planning and construction of programming assignments for teaching introductory computer programming. In R. D. Tennyson (Ed.), Automating instructional design, development, and delivery (NATO ASI Series F, Vol. 119, pp. 61–77). Berlin, Germany: Springer Verlag.