Abstract Title Page
Not included in page count.

Title: How to schedule multiple graphical representations? A classroom experiment with an intelligent tutoring system for fractions

Author(s): Rau, M. A., Aleven, V., & Rummel, N.

Abstract Body

Limit 4 pages single spaced.

Background:

Description of prior research and its intellectual context.

Graphical representations (GRs) of the learning content are often used for instruction (Ainsworth, 2006). When used in learning technology, GRs can be especially useful since they allow for interactions across representations that are physically impossible, for instance by dragging and dropping symbolic statements into a chart that automatically updates to display the information graphically (Moyer, Bolyard, & Spikell, 2002). However, learning with multiple GRs (MGRs) is challenging. An important pre-requisite for benefiting from the multiplicity of different GRs is that students conceptually understand each one of them (Ainsworth, 2006). Fractions are one of the many areas in mathematics in which multiple GRs are used extensively (National Advisory Board Panel, 2008). In a prior study, we found experimental evidence that students working with multiple GRs of fractions (e.g., circles, rectangles, and number lines) outperform students who work with a single GR (SGR), although only when prompted to explain how the GRs (e.g., half a circle) relate to the symbolic representation (e.g., 1/2) (Rau, Aleven, & Rummel, 2009). Our results demonstrate that understanding each of the GRs (i.e., by relating the GRs to the concepts of numerator and denominator) is essential learners' to benefit from MGRs.

Research Questions:

Description of the focus of the research.

When designing intelligent tutoring systems (ITS) that use MGRs, designers must decide how to temporally sequence the GRs. How often should the curriculum alternate between GRs? Practice schedules are likely to impact how students understand each GR. In particular, it may matter whether students learning tasks with the same GRs are practiced in a "blocked" manner (e.g., circle – circle – rectangle – rectangle) or are interleaved with practice of other GRs (e.g., circle – rectangle – circle – rectangle). Research on contextual interference shows that interleaving task types (i.e., sequences of tasks that vary in topics) leads to better learning results than blocking task types (Battig, 1972; de Croock, Van Merrienboer, & Paas, 1998). A common interpretation of this finding is that interleaved practice encourages deep processing (de Croock, et al., 1998). Since students cannot hold all relevant knowledge components in working memory, they must reactivate task-specific knowledge throughout the task sequence. When incorporating MGRs in an ITS, a relevant question is: Will students benefit most from blocked or interleaved MGRs?
In a prior study, we contrasted the effects of interleaving curricular topics while blocking MGRs and interleaving GRs while blocking topics. Our results demonstrated that the benefits from interleaving topics are larger than the benefits of interleaving GRs (Rau, Aleven, & Rummel, 2010). However, the question of whether interleaving GRs *in addition* to interleaving topics leads to better learning remains open. Our study was designed to addresses this question.
In addition implementing different practice schedules of MGRs, we included SGR control conditions in order to replicate our earlier finding that MGRs lead to higher learning gains than a SGR.

Setting:

Description of the research location.

Students participating in this study worked on an ITS developed for the purpose of the study. Students worked on the system as part of their regular math instruction. Specifically we developed several versions of an example-tracing tutor for fractions learning, using CTAT (Aleven, McLaren, Sewall, & Koedinger, 2008). This type of tutor behaves like a Cognitive Tutor (Koedinger & Corbett, 2006) but relies on examples of correct and incorrect solution paths rather than on a cognitive model. The design of the fractions tutor was informed by results from our previous studies (Rau, et al., 2009, 2010) as well as on small-scale user studies with 5th- and 6th-graders. All problems include interactive GRs (figure 1) and reflection prompts at the end of each problem (figure 2).

Participants:
Description of the participants in the study: who, how many, key features, or characteristics.

Five hundred and eighty-seven 4th- and 5th-grade students from six different schools (31 classes) participated in the study. We excluded students who missed at least one test day, and who completed less than 67% of all tutor problems (to ensure that students in the MGR conditions encountered all three GRs). This results in a total of N = 290 (n = 63 in blocked, n = 53 in moderate, n = 52 in fully interleaved, n = 62 in increased, n = 21 in single-circle, n = 20 in single-rectangle, n = 19 in single-number-line).

Intervention:
Description of the intervention, program, or practice, including details of administration and duration. For Track 2, this may include the development and validation of a measurement instrument.

All students worked with our tutor for fractions for about 5 hours at their own pace. The tutors used in the study included three interactive GRs of fractions: circles, rectangles, and number lines. The fractions tutor covered six task types: Identifying fractions from MGRs, making MGRs of symbolic fractions, reconstructing the unit from unit fractions, reconstructing the unit from proper fractions, identifying improper fractions from MGRs, and making MGRs of improper fractions.

Students solved each problem by interacting with both symbols and interactive GRs. Students manipulated the GRs in various ways: by clicking on fraction pieces to highlight them, by dragging-and-dropping fraction pieces, and through buttons to change the partitioning of the GRs. The tutor interfaces updated interactively after each step to show the next step to work on, and to emphasize parts of the GRs that were conceptually relevant for the subtask at hand through color-highlighting.

Students received error feedback and hints on all steps. Error feedback messages were designed to make students reconsider their answer using the MGRs, or by reminding them of a previously introduced principle. Hint messages provided conceptually oriented help, often in relation to the GR. Finally, reflection prompts were provided at the end of each tutor problem to help students reflect on the conceptual aspects demonstrated by the GRs. We found these prompts to be effective in an earlier experimental study (Rau, et al., 2009).

Research Design:
Description of the research design.

Table 1 illustrates the practice schedules of task types and GRs for the four MGR conditions. In all conditions, students worked through the same sequence of task types and fraction problems, and switched task types after 9 of a total of 108 problems. Each task type was revisited three times. This procedure corresponds to the most successful level of interleaving task types in our prior experiment (Rau, et al., 2010). We randomly assigned students to one of seven conditions. In the *blocked* condition, students switched GRs after 36 problems. In the *moderate* condition, students switched representations after every six problems. In the *fully interleaved* condition, students switched representations after each problem. In the *increased* condition, the length of the blocks was gradually reduced from twelve problems at the beginning to a single problem at the end. Finally, students in the three SGR conditions worked on all tutor problems with only the circle, the rectangle, or the number line, respectively.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.
For Track 2, this may include the use of existing datasets.

We assessed students' knowledge of fractions at three test times. Three equivalent test forms were created, and we randomized the order in which they were administered. The tests included four knowledge types. Area model problems (i.e., problems that involved circles and rectangles) and number line problems were considered to be knowledge reproduction, and fraction comparison, and proportional reasoning were included as transfer items. The theoretical structure of the test (i.e., the four knowledge types just mentioned) resulted from a factor analysis performed on the pretest data.

As mentioned, we analyzed the data of N = 290 students. There was no significant difference between conditions with respect to the number of students excluded ($\chi^2 < 1$). There were no significant differences between conditions at pretest for any dependent measure, ps > .10. We used a hierarchical linear model (HLM, see Raudenbush & Bryk, 2002) with four nested levels to analyze the data. We modeled performance for each of the three tests for each student (level 1), differences between students nested within classes (level 2), differences between classes nested within schools (level 3), and differences between schools (level 4).

More specifically, the following HLM model was fitted to the data:

$\text{score}_{ij} = \text{test}_j + \text{condition}_i + \text{test}_j\text{*condition}_i + \text{preScore}_i + \text{preScore}_i\text{*condition}_i + \text{numProblems}_i + \text{student(class)}_i + \text{class(school)}_i + \text{school}_i$;

with the dependent variable score_{ij} being student_i's score on the dependent measures at test_j (i.e., immediate or delayed posttest).

In order to analyze whether students with different levels of prior knowledge benefit differently from our conditions, we included students' pretest scores as a covariate (preScore_i), and modeled the interaction of pretest score with condition ($\text{preScore}_i\text{*condition}_i$).

To analyze learning gains, used a modified model in which we included pretest score in the dependent variable, instead as a covariate. In addition, we used planned contrasts and post-hoc comparisons to clarify results from the HLM analysis. All reported p-values were adjusted using the Bonferroni correction.

Results:

Description of the main findings with specific details.

Learning effects. The main effect of test was significant for number line test items ($p < .01$), fraction comparison ($p < .05$), and proportional reasoning ($p < .01$). The interaction between test and condition was significant for area model test items ($p < .05$). These results show that students (regardless of condition) improved on number line, fraction comparison, and proportional reasoning items. On area model items, students' learning gains depended on the condition. Post-hoc comparisons comparing students' scores at the immediate posttest and the delayed posttest compared to the pretest showed significant learning gains at the delayed posttest for most of the MGR conditions on area models, number line, and proportional reasoning items. On comparison items, only the moderate condition demonstrated significant learning gains at the delayed posttest. Finally, we found no significant learning gains for the SGR conditions except for the single-circle condition at the delayed posttest on proportional reasoning.

Scheduling of MGRs. There was no significant main effect of condition on any posttest scale, indicating that there was no global effect of the practice schedules of MGRs across the posttests. The interaction between test time and condition for area model items and fraction comparison items was significant (ps $< .05$), indicating that the effect of practice schedules depends on test time. The interaction between pretest score and condition was marginally significant for proportional reasoning items ($p < .10$), demonstrating that students with different pretest scores benefit from different practice schedules. Post-hoc comparisons for the immediate and delayed posttest demonstrated an advantage of one of the interleaved conditions (i.e., fully interleaved, moderately interleaved, or increasingly interleaved) over the remaining conditions.

To clarify the interaction between pretest score and condition on proportional reasoning items, we computed post-hoc comparisons for students with extremely low or high pretest scores. For students with a pretest score of 15%, 20%, and 25%, we found a significant advantage for the fully interleaved over the blocked condition (ps $< .05$). We found no differences for high prior knowledge students.

Effects of MGRs vs. SGRs. Planned contrasts comparing the MGR conditions to the SGR control conditions showed a significant advantage for the MGR conditions over the SGR conditions for number line test items at delayed posttest ($p < .05$).

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

Our results demonstrate significant learning gains for students who worked with a tutoring system that supports learning with MGRs of fractions. The gains persist until one week after the study when we administered the delayed posttest. These learning gains are consistent for students in the MGR conditions on all posttest scales but fraction comparison. The fact that students' performance on fraction comparison does not improve may be due to the fact that this topic was not the focus of the tutor. The lack of learning gains in the SGR conditions demonstrates that the learning gains in the other conditions are not merely due to learning of the test. Further support for learning with MGRs comes from the finding that the MGR conditions outperformed the SGR conditions on number line items at the delayed posttest.

In addition, we found limited evidence that the practice schedule of MGRs matters. Interleaving MGRs (to a full extent or in an increased fashion) is beneficial, albeit we could so far only establish this effect on a subset of measures. We therefore carefully conclude that designers of ITS should employ an interleaved practice schedule of MGRs, in order to encourage students to engage in deep processing of GR-specific aspects of the learning content.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16(3), 183-198.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2008). Example-tracing tutors: A new paradigm for intelligent tutoring systems. International Journal of Artificial Intelligence in Education, 19(2), 105-154.

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Vos (Eds.), Levels of processing in human memory (pp. 131-159). Hillsdale, NJ: Lawrence Erlbaum.

de Croock, M. B. M., Van Merrienboer, J. J. G., & Paas, F. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills: Effects on transfer performance and invested mental effort. Computers in Human Behavior, 14(2), 249-267.

Koedinger, K. R., & Corbett, A. (2006). Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom. New York, NY, US: Cambridge University Press.

Moyer, P., Bolyard, J., & Spikell, M. A. (2002). What are virtual manipulatives? Teaching children mathematics, 8, 372-377.

Panel, N. M. A. (2008). Foundations for Success: Report of the National Mathematics Advisory Board Panel: U.S. Government Printing Office.

Rau, M. A., Aleven, V., & Rummel, N. (2009). Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. Paper presented at the 14th International Conference on Artificial Intelligence in Education Brighton, UK.

Rau, M. A., Aleven, V., & Rummel, N. (2010). Blocked versus Interleaved Practice with Multiple Representations in an Intelligent Tutoring System for Fractions Paper presented at the 10th International Conference of Intelligent Tutoring Systems.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods (2nd ed.). Newbury Park, CA: Sage Publications.

Appendix B. Tables and Figures

Not included in page count.

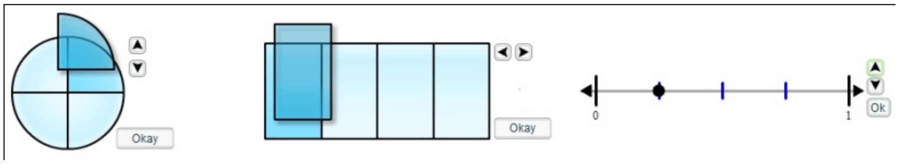

Figure 1. Interactive representations used in fractions tutor: circle, rectangle, and number line.

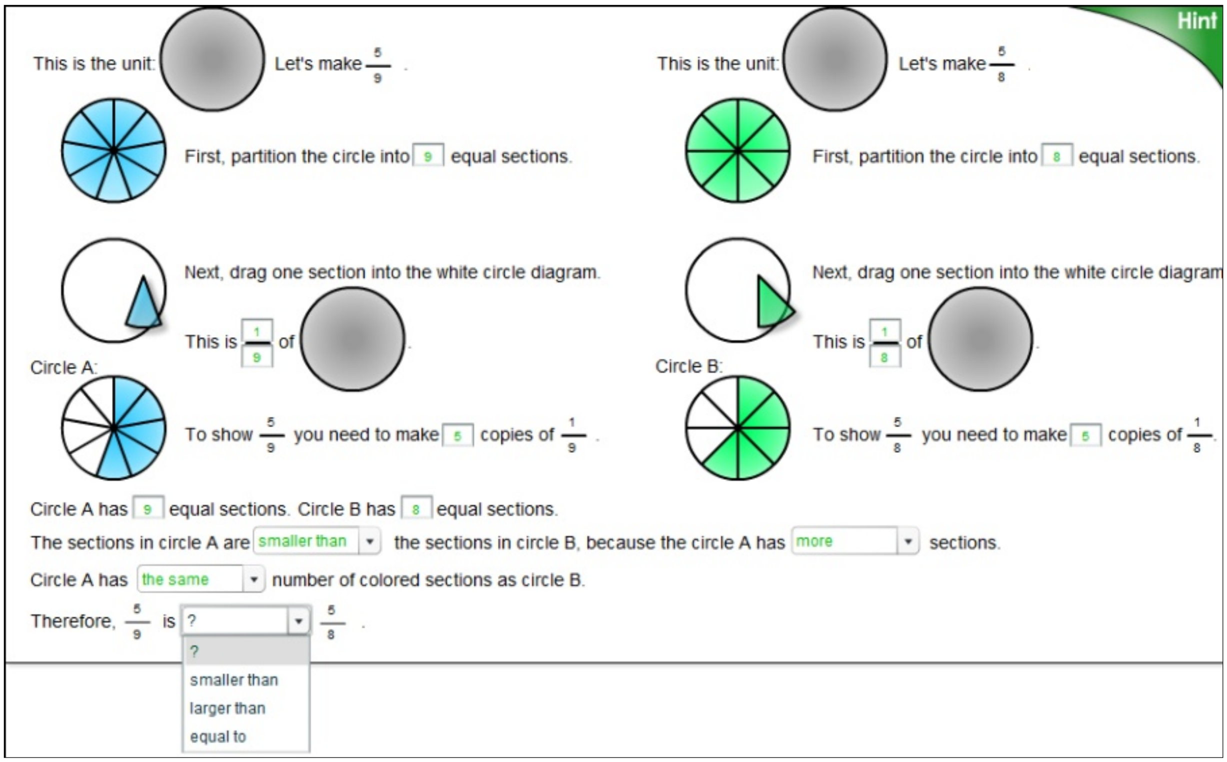

Figure 2. Making a circle given a symbolic fraction, combined with prompts to compare the two fractions. Reflection prompts are implemented with drop-down menus shown at the bottom.

Table 1. Practice schedule for MGR conditions for all six task types (TT). Each TT was revisited three times. Letters indicates a tutor problem and its GR: circle (C), rectangle (R), or number line (N).

TT	Blocked	Moderate	Fully Interleaved	Increased
1	c–c–c–c–c–c–c–c–c	c–c–c–c–c–c–r–r–r	c–r–n–c–r–n–c–r–n	c–c–c–c–c–c–c–c–c
2	c–c–c–c–c–c–c–c–c	r–r–r–n–n–n n–n–n	c–r–n–c–r–n–c–r–n	c–c–c–c–c–c–c–c–c
3	c–c–c–c–c–c–c–c–c	c–c–c–c–c–c–r–r–r	c–r–n–c–r–n–c–r–n	r–r–r–r–r–r–r–r–r
4	c–c–c–c–c–c–c–c–c	r–r–r–n–n–n n–n–n	c–r–n–c–r–n–c–r–n	r–r–r–r–r–r–r–r–r
5	c–c–c–c–c–c–c–c–c	c–c–c–c–c–c–r–r–r	c–r–n–c–r–n–c–r–n	n–n–n–n–n–n–n–n–n
6	c–c–c–c–c–c–c–c–c	r–r–r–n–n–n n–n–n	c–r–n–c–r–n–c–r–n	n–n–n–n–n–n–n–n–n
1	r–r–r–r–r–r–r–r–r	r–r–r–r–r–r–n–n–n	r–n–c–r–n–c–r–n–c	r–r–r–n–n–n–c–c–c
2	r–r–r–r–r–r–r–r–r	n–n–n–c–c–c c–c–c	r–n–c–r–n–c–r–n–c	r–r–r–n–n–n–c–c–c
…	…	...	…	…
1	n–n–n–n–n–n–n–n–n	n-n–n–n–n–n–c–c–c	n–c–r–n–c–r–n–c–r	n–c–r–n–c–r–n–c–r
2	n–n–n–n–n–n–n–n–n	c–c–c–r–r–r r–r–r	n–c–r–n–c–r–n–c–r	n–c–r–n–c–r–n–c–r
…	…	...	…	…