Predictors of Transfer of Experimental Design Skills in Elementary and Middle School Children[1]

Stephanie Siler[1], David Klahr[1], Cressida Magaro[1], Kevin Willows[1], and Dana Mowery[2]

[1]Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue, 15213, Pittsburgh, PA, United States
[2]Pittsburgh Science and Technology Academy, 107 Thackeray St., 15213, Pittsburgh, PA, United States
[1]{siler, klahr, cmagaro, KevinWillows} @cmu.edu
[2]dmowery1@pghboe.net

Abstract. A vital goal of instruction is to enable learners to transfer acquired knowledge to appropriate future situations. For elementary school children in middle-high-SES schools, "explicit" instruction on the Control of Variables Strategy (CVS) has proven to be very effective at promoting transfer, even after time delays, when administered by human instructors [1], [2] and when administered by our computer tutor ("TED" for Training in Experimental Design). However, when the same instruction was delivered to students in low-SES schools, near—but especially far—transfer rates were lower. We discuss our findings of the predictors of transfer in this population, and an initial investigation assessing the causal status of one candidate factor for far transfer, understanding the logic of CVS. Finally, we discuss the potential implications of these findings for ways to adapt instruction to individual students.

Keywords: transfer; experimental design skills; computer-based tutor

1 Introduction

The primary goal of instruction is to enable learners of widely varying abilities to transfer newly acquired knowledge to future situations. Over the past decade, our lab has studied the effects of several instructional and contextual factors, as well as student characteristics, on their ability to learn the core procedural and conceptual knowledge elements associated with simple experimental design and to transfer that knowledge to appropriate future contexts [1], [2], [3]. The "Control of Variables Strategy" (CVS) is an important domain-general topic in elementary and middle school science. It involves controlling all variables in an experiment except for the

[1] The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H060034 to CMU and the National Science Foundation through the Pittsburgh Science of Learning Center, Grant SBE0354420. The opinions expressed are those of the authors and do not represent views of IES or the U.S. Department of Education.

focal variable, i.e. the variable whose effect on an outcome is being investigated. In this paper, we summarize our previous work in order to lay the groundwork for the (ongoing) creation of an adaptive, intelligent tutor to teach CVS. Our primary effort is to identify student-specific factors that reliably predict different levels of transfer, and then to devise ways to diagnose and rules for the tutor to respond to those individual factors in teaching children about CVS.

Our prior studies have revealed a consistent relationship between the average SES of the students in different participating classrooms and their ability to learn CVS. (Throughout this paper, our proxy for SES is the proportion of students in a school eligible for free or reduced-price meals.) In middle- and high-SES classrooms, human-delivered explicit CVS instruction that draws students' attention to the reasons why an unconfounded experiment is, in fact, unconfounded has been very effective in promoting not only *near transfer* (i.e., successfully designing experiments in the domain used during instruction) but also *far-transfer* (i.e., successfully designing and evaluating experiments in domains other than the instructional domain) [1], [2], [3]. However, in low-SES classrooms, where students typically have had less exposure to science inquiry, and have poorer reading and math skills — this type of instruction has been much less successful [4]. One of the aims of our current research is to determine what *cognitive* factors – probably correlated with the distal SES measure – are actually influencing the differential effectiveness of our instruction for high and low SES students.

Over the past several years, we have been incrementally converting this explicit instruction into an adaptive, intelligent, tutor that will assess student knowledge and provide highly tailored instruction on the procedural and conceptual aspects of CVS. We call our system the "TED" (Training in Experimental Design) tutor. In its current non-adaptive state, the TED tutor – running in either stand-alone or networked contexts – consists of the following series of components that closely replicate the procedures used in our "human teacher" experimental training studies [1], [2], [3] as well as several human-teacher classroom implementations [4], [5].

1. <u>Story pretest</u>: 6 items requiring students to design (3 items) or evaluate (3 items) experiments presented as "story problems," and to provide a rationale for their responses. These story problems include three different contexts—cookie baking, drink sales, and rocket ship design. For each context, students are first asked to design an experiment to test a particular variable by selecting values for each of three variables in two conditions. Then they explain why they set up the experiment as they did. Following the design question, students evaluate a given experiment in that context as a "good" or "bad" way to find out whether the focal variable makes a difference and explain their response. If they indicate it is a bad way, they were asked to change it into a good experiment, which required them to change at least one value setting.
2. <u>Ramps intro</u>: A Flash-based section (~ 2 min) presenting simple animated color line drawings – accompanied by audio voice-over and dynamic visual pointers that introduce the four variables relevant in the ramps apparatus (steepness of slope, length of run, surface type, and ball type).
3. <u>Ramps pretest</u>: Similar visual and audio presentation of 4 test items. Each item requires students to (a) design an experiment to determine the causal role of one of the four ramp variables and (b) provide a rationale for their design. Students select one of two values to use for each of four factors on each of two comparison ramps by moving parts of the diagram or selecting values in a table. The diagram and table are linked (e.g., a text

selection of a "steep" ramp automatically raises the ramp steepness in the diagram, or increasing the ramp steepness in the diagram automatically generates "steep" in the corresponding table cell). At present, student justifications are entered as free text, but we are currently replacing text boxes with drop-down menus for students to select rationales for their designs. This will greatly simplify the challenge of using free-form explanations to infer knowledge states in the adaptive version of TED.

4. Introductory video: A brief (~ 2 min) video that consists of a professional actress presenting an instructional "lecture," consisting of an introduction to experimental design, its purpose, scope, and the central idea of "comparing and contrasting" things to find out whether or not they produce different outcomes, accompanied by simple animated line drawings supported by Flash-based graphics. (Note: In later versions of TED, this video will precede the ramps intro.)
5. Explicit Instruction "EI": Based on "explicit" CVS instruction developed in our previous studies [1], [2], [3], in this portion of instruction – delivered in the same format as the ramps pretest – students are presented with three different pairs of contrasting ramps set-ups (i.e., "experiments"). For each, they are probed for whether and why the design is or is not "a good way" to find out about the focal variable; and whether the design would allow them to "know for sure" whether the focal variable made a difference in outcomes; and why or why not. Students type responses to the "why" questions into text boxes. After responding to these deep questions, they are given feedback and an explanation for why the design could—or could not—lead to valid inferences about the focal variable. Any experimental confounds are corrected by the tutor, and students answer the same two deep probes described above. Students are then given feedback and an explanation for why the unconfounded experiment would lead to valid inferences about the focal variable.
6. Ramps posttest: Identical to the ramps pretest; this is our measure of *near-transfer* performance, completed by students immediately after the Explicit Instruction ("EI").
7. Story posttest (identical to the story pretest); this is our measure of *immediate far-transfer* performance, completed by students the following day.
8. Delayed story posttest (identical to story pre/posttest, except for different focal variables); this assesses students' delayed far-transfer performance three weeks later.

In the version of the TED tutor as of this writing, students progress from one component to the next in the sequence given above. We are presently integrating Bayesian knowledge tracing, driven by students' menu-based responses on the ramps pretest, to decide when to take students through alternative assessment and instructional paths. At minimum, the adaptive version of TED will assess three CVS procedural knowledge components or "rules": R1: identify the focal variable given in the problem statement; R2: contrast levels of the focal variable across conditions; R3: control all other variables. Based on their responses, students may be given instruction in one or more of these rules prior to entering the EI phase.

In a recent evaluation, the TED tutor was compared to human instructors who followed the same script and procedure (described above) but used physical rather than virtual materials. Because there was no difference between TED and human instructors in near and far transfer outcomes in either population of students who did not display CVS mastery on the story pretest, human and TED-tutored students were combined. As shown in Table 1, and consistent with previous findings [4], the mastery rates of two classrooms of low-SES 5^{th}-grade children (L2, n = 16; L3, n =

14) were lower than those of their middle-SES counterparts[2], (n = 50) particularly on the far-transfer assessments, where the transfer mastery rates were more than four times greater in the middle-SES classroom.

Table 1. Summary of transfer mastery rates from a recent TED evaluation.

	Mastery ramps pretest	Near transfer mastery (ramps post)[a]	Immediate far transfer[b]	Delayed far transfer[b]
Mid-SES	20%	87.5%	61.9[3]%	62.5%
Low-SES	10%	60.0%	13.4%	15.4%

[a] At least 3 of 4 CVS set-ups. [b] At least 5 out of 6 CVS set-ups.

2 Predictors of CVS transfer

In what follows, we first identify the predictors of these near and far transfer outcomes for the low-SES student population. Then we look at the relationship between student-specific measures and these predictors. Based on these findings, we propose ways to make instruction in the TED tutor adaptive to individual student users, in part by informing the Bayesian model of key knowledge components that are strong predictors of learning and transfer. We used data from the previously described study (classes L2 and L3) as well as from an earlier evaluation performed as part of the TED project, in which a science teacher administered EI to 6th-grade students in a low-SES classroom (L1, n = 23).

We first looked at which initial knowledge and standardized measures were most highly correlated with posttest performance for the two low-SES 5th-grade classrooms. These measures were ramps pretest, story pretest, standardized (CTB/TerraNova[4]) reading comprehension, science, nonverbal "IQ," and verbal (or deductive) reasoning national percentile scores. The verbal deductive reasoning measure on the CTB/TerraNova test assesses the skill of identifying a conclusion that is based only on information given. To recognize the correct response, students must integrate the information to produce that response. This task also requires that students do *not* select distractors that may be consistent with common knowledge. The following is an example of a verbal deductive reasoning practice item (Level 2, for U.S. grades 4-5), with the correct response italicized:

A fire must have heat, air, and fuel or it will not burn.
Wood can be used as fuel for a fire.
The scouts made a campfire.
(a) The scouts used wood to make their campfire.

---

[2] The two low-SES classes were from schools in which 95% and 59% of students were eligible for free or reduced lunch; the middle-SES class was from a school where 20% of students were eligible.

[3] Due to a discrepancy in the format of the computer-administered story posttest (later modified), only human-tutored students are included for immediate far transfer mastery rates; students in both conditions are included for the delayed far transfer mastery rate.

[4] For more information on this test, go to: http://www.ctb.com/mktg/terranova/tn_technical.jsp

(b) The scouts toasted marshmallows over their fire.
(c) The campfire had heat, air, and fuel.
(d) The campfire burned for a long time.

Predictors of near transfer. Of these measures, only reading comprehension was significantly related to near-transfer performance (r = +.47, p = .03) in a forward regression. This relationship did not differ by classroom (L2 or L3) or condition (Human- or TED-tutored). The same result was found in the earlier evaluation, in which a science teacher administered EI to 6^{th}-graders in a low-SES classroom (L1). Using the same variables in a forward regression, only reading comprehension was significantly related to ramps posttest scores (r = +.87, p < .001). In both cases, because instruction was presented orally by the teacher or with audio voice-over in TED, we believe that a more general comprehension skill may underlie the relationship between reading comprehension and near transfer than reading ability per se. These results are summarized in Table 2.

Table 2. Assessment and standardized test correlates of transfer performance.

Assessment	(L1) (classroom instruction)	(L2 & L3) (Human & TED)
Near transfer (ramps post)	Reading comprehension	Reading comprehension
Immediate far transfer	(n/a)	Deductive reasoning
Delayed far transfer	Deductive reasoning	Story post & Deductive reasoning

Predictors of far transfer. Of both pretest and all standardized measures (reading comprehension, science, nonverbal "IQ," and verbal/deductive reasoning), only deductive reasoning was significantly related to the measure of immediate far transfer (r = +.58, p = .006). Similarly, when immediate story posttest was also included in a forward regression, only deductive reasoning and immediate story posttest were significantly related to delayed story posttest score (r = +.49, p = .04, for both variables). Likewise, in L1, including all these independent variables (with the exception of the immediate story posttest, not administered for L1), only deductive reasoning was significantly related to the delayed story-evaluation posttest (r = +.82, p = .001).

Deductive reasoning may play a role during learning of CVS that may account for far transfer. Prior research [6] has found that conceptually-oriented explanations are predictive of procedural transfer and higher quality explanations were positively related to performance. Additionally, Kuhn and Dean [7] speculated that helping students to understand why to use CVS is critical metastrategic knowledge necessary for transfer. Therefore, we performed a finer-grained analysis and coded for students' "highest quality" responses—those that demonstrated a complete understanding of the determinate nature of an unconfounded set-up (or the indeterminate nature of a confounded experiment), or "CVS logic." For example, when given the probe: "Imagine the balls rolled different distances. Could you tell for sure that the surfaces caused the difference?," one TED-tutored student responded: "Yes. Because everything is the same and if there is a difference it's because of the surface." This response explicitly demonstrates an understanding of the determinate causal link between the focal variable and outcome. In contrast, the following response to the same probe, though correct, is of lower quality because it does not explicitly express the causal link between the variable and outcome differences: "Yes. Because everything is the same except [surface]."

Deductive reasoning—as assessed in the TerraNova—could be related to the quality of explanation because both involve integrating and drawing conclusions from given information. Whether or not students explicitly expressed this causal logic during the experimental evaluation portion of instruction was more highly related to their deductive reasoning scores than any other pretest, standardized measure, or correct responses to questions posed during the "EI" component. Furthermore, deductive reasoning was more highly related to an expression of causal logic than other coded measures. Thus, expression of CVS logic and deductive reasoning appear to be highly inter-related.

Regarding near transfer performance, when reading comprehension and ramps pretest scores were included in the regression, there was no correlation between student expression of the logic of CVS and ramps posttest. Thus, this deeper understanding did *not* predict near transfer performance. However, with both deductive reasoning and expression of the causal logic in the regression model, only whether students expressed CVS logic during the experimental evaluation phase was significantly related to immediate far transfer performance. These correlations are shown below:

Deductive reasoning → CVS logic → Immediate & Delayed story posttest

Thus, CVS logic understanding—assessed during instruction—may explain the relationship between deductive reasoning and far transfer. Similarly, in a regression with immediate story posttest and deductive reasoning as covariates, delayed far transfer was only predicted by whether or not students explicitly expressed CVS logic, and not by immediate story posttest performance or deductive reasoning. Nor did it interact with condition, and thus was predictive of far transfer performance for both human- and TED-tutored students as anticipated. Thus, again this measure of deep conceptual understanding was predictive of far transfer performance, and may play a causal role in it. Note that if knowledge integration is primarily related to far transfer, this may explain why self-explanation prompts did not improve near transfer CVS performance in a recent study [8].

3 Pilot Study

For an initial test of whether CVS logic understanding improves far transfer performance, we compared students given "basic" TED-delivered CVS instruction to students who were additionally prompted to think about the link between the experimental set-up and the conclusions that could be drawn about causality (i.e., the logic of CVS)[5]. If this deep understanding is related to far transfer, then students

[5] This was done as part of a larger study comparing TED instruction to a control lesson on CVS.

given the added prompt should be more likely to express CVS logic and out-perform control students on the story posttests.

2.1 Participants, Design, and Procedure

Participants were 8th-grade students in one science class at a local magnet school participated in this pilot study. The majority of students in this school (69%) were eligible for free or reduced-price lunch; thus, we consider this a low-SES population. Of the 22 students who completed the pretest, after removing from analyses students who: (a) did not have parental permission for data use, (b) showed incoming CVS mastery, (c) did not have available reading levels, and (d) were absent for some of the instruction, only 11 students remained. Given the small sample, the results presented below, although quite interesting, are only suggestive at this point.

Students were randomly assigned to the "baseline" or "added-questions" (AQ) condition in a two-condition, between-subjects design. Each phase of the procedure took place during the final period of the school day, their regular science class. The procedural sequence is the same as described in the introduction. On the first day, all students completed the computerized story pretest (described earlier). The EI phase was split between the second and third days. On the second day, students evaluated the first (unconfounded) experiment and received explicit instruction on experimental design. On the third day, students completed the EI phase by evaluating and receiving explicit instruction on the second and third experimental designs, both confounded. Thus, students in both conditions evaluated the same number of experiments.

The instructional script of the EI phase for the first experimental evaluation is shown in Table 3. Students in both conditions received the same questioning until after Q4. At that point, students in the baseline condition were given a procedural explanation for why the unconfounded design was good. Students in the AQ condition were prompted to identify potential causal factors in the set-up and given feedback on their response. Then all students were told why the experiment would allow them to determine whether the focal variable was causal (i.e., the logic of CVS).

After the EI phase on the third day, students viewed a brief summary of the lesson and first completed the ramps posttest, then the story posttest and answered four standardized questions on paper. Finally, three weeks later, students completed a paper version of the delayed story posttest, followed by the four standardized CVS test items. The delayed story posttest was identical to the story pretest and immediate posttest, but targeted different focal variables. (Due to space constraints, we will not discuss the results of the standardized items in detail here, other than to note that there were no significant differences between conditions on this measure).

Table 3. One evaluation cycle of explicit instruction “EI” phase by condition [a].

Baseline	Added-questions (AQ)
Exp 1: Unconfounded experiment (focal variable = surface)	
Q1: “Is this experiment a good way to find out whether the balls go different distances just because of the ramp surface?” (Y/N) Q2: “Why or why not?” (typed response) Q3: “Imagine the balls rolled different distances. Could you tell for sure that the surfaces caused the difference?” (Y/N) Q4: “Why or why not?” (A3 Feedback: You’re right/Actually, we could tell for sure from this comparison whether changing the surface (or making the surfaces different) causes a change in how far the balls roll.)	
E5: “The reason we could tell for sure is that the only thing different between these two ramps is the surface. One is [value 1] and the other is [value 2]. The ramps are built exactly the same way, except for the surface.”	**Q5: “What else besides the [focal variable] could have made the balls roll different distances?”** *(student selects one or more variables)* (Feedback on A5: “Right/Actually, ONLY the different surfaces could have caused one ball to roll farther than the other, because only the surfaces are different between the two ramps.”)
E6 (CVS logic explanation): “The ONLY thing that is different is the thing Amal is trying to find out about. Everything else is the same. They have the same slope, the same ball, and the same starting position. If one of the balls rolled farther, Amal would know that it could only be the surface that caused this result, since it’s the only thing different between the two ramps. Amal could say whether the surface affects how far the balls roll. So, Amal made a GOOD experiment!”	

[a] This cycle was repeated twice for each of two initially confounded experiments—once for confounded state and once for fixed unconfounded state, for a total of 5 added questions.

4 Results

Understanding of CVS logic. Students’ responses during the EI phase of the intervention were coded for understanding of CVS logic, that is, whether they included expressions of the causal indeterminacy of counfounded experiments or the determinacy of unconfounded experiments. Counter to expectation, students in the AQ condition were no more likely to give at least one CVS logic statement in the EI phase than students in the baseline condition (4 of 5, and 3 of 6, respectively), Fisher’s exact $p = .55$. Nor were they more likely to justify their designs on the ramps posttest in terms of causal logic: no student in either condition did so.

Near and far transfer. As expected, the majority of students (80% and 89% in baseline and AQ conditions, respectively) achieved near-transfer mastery. These rates did not differ and were similar to those of middle-SES 5th-graders (Table 1).

Students who were asked the additional questions in the TED tutor did not significantly out-perform students in the baseline condition on the immediate story posttest, $F(1, 8) = 0.16$, $p = .70$ (Table 4). Because the prior period ran late, students had less time than they may otherwise have taken on the immediate story posttest. Students in the AQ condition may have been even more rushed: they spent less time on the immediate story posttest than students in the baseline condition (M = 5.97 min, SD = 2.26; M = 9.03 min, SD = 3.21, respectively), $F(1, 10) = 3.80$, $p = .08$. With time on immediate posttest included in ANCOVA, there was a nearly significant condition by reading level interaction, $F(1, 6) = 5.80$, $p = .05$. Lower-reading students (basic and below basic) tended to perform better in the AQ condition whereas the higher-reading students (proficient and advanced) tended to perform better in the baseline condition. This suggests adapting TED instruction to students' reading level by assigning lower-reading students to the AQ version and higher-reading students the baseline version of EI.

Table 4. Story test means (and standard deviations) and mastery rates by time.

Condition	Immediate	Delayed	Immediate mastery rate [a]	Delayed mastery rate [a]
Added-questions	4.17 (1.33)	5.17 (1.60)	33%	67%
Baseline	3.00 (2.45)	2.75 (2.75)	40%	25%

[a] At least 5 out of 6 CVS set-ups.

With respect to delayed far transfer performance, though students in the AQ condition tended to score higher than those in the baseline condition (Table 4), this difference missed significance, $F(1, 7) = 3.44$, $p = .11$. Failure to reach significance may be due to small sample size. However, the AQ students showed significantly higher story test gains from the immediate to delayed posttest, $F(1, 11) = 5.17$, $p = .04$, where only students who answered the added questions showed significant immediate to delayed posttest gains. In sum, though students in the AQ condition were no more likely to express the logic of CVS during the EI, they tended to perform better on the far transfer assessments. Because this result does not support the hypothesis that CVS logic understanding causes far transfer, we sought to determine which factors were predictive of far transfer.

Predictors of far transfer. The factors we investigated in pair-wise correlations were story pretest, ramps pretest score, expression of CVS logic, the number of correct responses to Q1 in Table 3 ("Is this a good way…") and Q3 ("Can you tell for sure…"), and reading level. No other standardized measures (e.g., deductive reasoning) were available.

In the baseline condition, reading level and expression of CVS logic were significantly related to immediate far transfer performance (r = +.88, $p = .048$; r = +.95, $p = .02$, respectively). When both expression of CVS logic and reading were included in a backward (or forward) step-wise regression, only expression of CVS logic remained in the model. However, no other factors were predictive of expression of CVS logic. Only immediate posttest score was significantly related to performance on the delayed story posttest (r = +.99, p = .01). These correlational links for students in the baseline condition, used to derive adaptive rules in the TED tutor, are shown below:

CVS logic → Immediate story posttest → Delayed story posttest

In the AQ condition, using the same variables as above but also including the number of correct responses to the added questions, neither expression of CVS logic nor reading level was related to far transfer performance. Rather, the number of correct responses to Q1 & Q3 (Table 3) was the best predictor of immediate far transfer performance (r = +.87, p = .002). In turn, reading level was the best—and only significant—predictor of number of correct Q1 and Q3 responses (r = +.95, p = .004). For the delayed story posttest, the number of correct responses on Q1 and Q3, the number of correct responses on the added questions, and the immediate story posttest were all significantly correlated with delayed story posttest performance (r = +.93, p < .001; r = +.88, p = .002; r = +.82, p = .006; r = +.92, p = .001, respectively). In both a forward and backward step-wise regression with these variables, only the number of correct responses on Q1 and Q3 remained in the model. These correlations for students in the AQ condition are shown below:

Reading → Correct responses on Q1 & Q3 → Immediate & Delayed story posttest

The number of correct responses to Q1 and Q3 was only significantly related to immediate and delayed far transfer performance for students in the AQ condition. It may be that correct responses to these items indicate a deeper understanding for students in the AQ than in the baseline condition (and incorrect responses may indicate greater confusion). However, the significant relationship between reading and number of correct responses to Q1 and Q3 questions may indicate that students with poorer comprehension skills do not understand the content of the EI as well as they might.

5 Implications for the TED tutor

In previous studies with low-SES 5th and 6th-grade students, reading comprehension was the best predictor of near transfer performance. This relationship was not found for the 8th-graders, likely because they were near ceiling on the ramps posttest. We believe that one way to address the needs of students with poorer comprehension skills is to reduce cognitive load by reducing the amount of information students must process in a given conversational turn. Previous work [9] found that shorter tutor turns were related to better posttest performance, especially for students with poorer comprehension skills. Consistently, in our earlier work that included human tutoring of children who failed to learn CVS from the "EI" phase, we found that presenting the rules of CVS in a more incremental way—while still emphasizing the rationales for applying them—often helped students to develop a robust understanding of CVS. Thus, students with lower reading comprehension scores may be given more incrementally-delivered training on controlling non-focal variables before entering the “EI” phase of the tutor (“Rule 3 training” in Fig. 1). In this training, students are asked to select values for one non-focal variable at a time and receive immediate procedural and conceptual feedback on their responses.

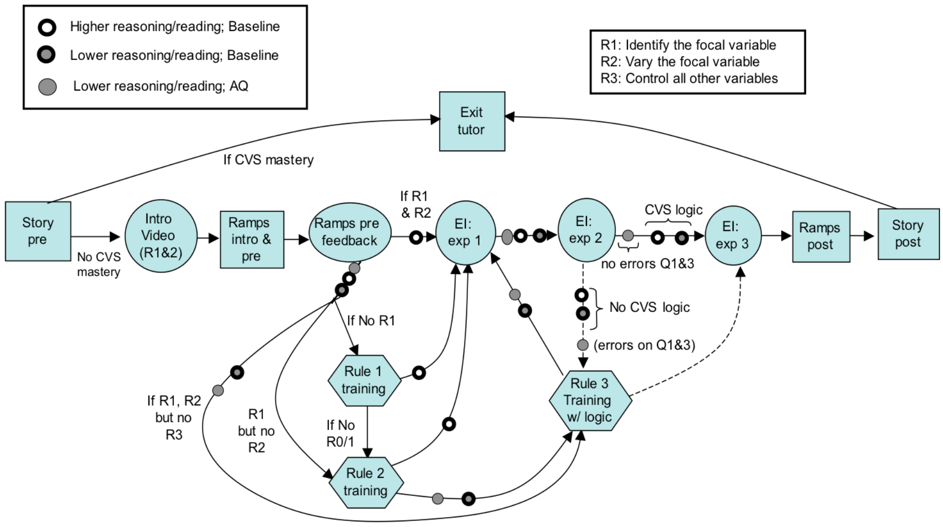

Figure 1. Adaptive pathways with current and future components of the TED tutor.

If the Bayesian knowledge-tracing model detects that a student likely lacks R1 or R2, remedial training can be given on these concepts (Fig. 1). Students with better reasoning or comprehension skills can then be ushered into the (baseline version of) the EI, whereas students with poorer skills can first be provided with the more incremental Rule 3 training prior to entering the EI phase. Students with poorer reading/reasoning skills, who tended to perform better in the AQ condition, may then go to this version of the EI.

As found in earlier studies, expression of CVS logic was the best predictor of far transfer performance for students in the baseline condition. Thus, students in this version who do not express the logic of CVS by the end of the second experimental evaluation can be directed into the R3 training module to ensure they understand why to control.

In the AQ condition, the number of correct responses to the Q1 and Q3 questions was predictive of far transfer performance. It is possible that students in the AQ condition of the pilot study were more likely to understand the logic of CVS yet failed to express it in their responses during EI, perhaps because they thought this concept was more obvious than students in the baseline condition. Students who answer these questions incorrectly may be diverted to R3 training (Fig. 1).

Evidence of the effectiveness of these modifications would include both improved near transfer performance and a weaker relationship between reading comprehension and near transfer performance than was found in the past. It would also include a greater proportion of students in the baseline version expressing an understanding of the logic of CVS, and in the AQ version, a weaker link between reading and correct responses to the EI evaluation questions and more correct responses. But of course, the ultimate test of the effectiveness of the chosen adaptations is whether a greater percentage of students of varying backgrounds and characteristics develops a robust concept of CVS.

References

1. Chen, Z., Klahr, D.: All Other Things Being Equal: Children's Acquisition of the Control of Variables Strategy. Child Dev. 70(5), 1098–1120 (1999)
2. Strand-Cary, M., Klahr, D.: Developing Elementary Science Skills: Instructional Effectiveness and Path Independence. Cog. Dev. 23(4), 488-511 (2008)
3. Klahr, D., Nigam, M.: The Equivalence of Learning Paths in Early Science Instruction: Effects of Direct Instruction and Discovery Learning. Psych. Sci. 15(10), 661-667 (2004)
4. Klahr, D., Li, J.: Cognitive Research and Elementary Science Instruction: From the Laboratory, to the Classroom, and Back. J. of Sci. Ed. and Tech. 4(2), 217-238 (2005)
5. Toth, E., Klahr, D. & Chen, Z.: Bridging Research and Practice: A Cognitively-based Classroom Intervention for Teaching Experimentation Skills to Elementary School Children. Cog. & Instr. 18 (4), 423-459 (2000)
6. Matthews, P., Rittle-Johnson, B.: In Pursuit of Knowledge: Comparing Self-explanations, Concepts, and Procedures as Pedagogical Tools. J. of Exp. Child Psych. 104(1), 1-21 (2009)
7. Kuhn, D., Dean Jr., D.: Is Developing Scientific Thinking All About Learning to Control Variables? Psych. Sci. 16(11), 866-870 (2005)
8. Sao Pedro, M., Gobert, J., Heffernan, N., Beck, J.: Comparing Pedagogical Approaches for Teaching the Control of Variables Strategy. In: Taatgen, N.A., van Rijn, H. (eds.) Proceedings of the 31st Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin (2009)
9. Siler, S., VanLehn, K.: Learning, Interactional, and Motivational Outcomes in One-to-one Synchronous Computer-mediated Versus Face-to-face Tutoring. Inter. J. of Artif. Intell. in Ed. 19(1), 73-102 (2009)