

Reasoning Mind Genie 2: An Intelligent Tutoring System as a Vehicle for International Transfer of Instructional Methods in Mathematics

George A. Khachatryan · Andrey V. Romashov ·
Alexander R. Khachatryan · Steven J. Gaudino ·
Julia M. Khachatryan · Konstantin R. Guarian ·
Nataliya V. Yufa

Published online: 10 June 2014

© International Artificial Intelligence in Education Society 2014

Abstract Effective mathematics teachers have a large body of professional knowledge, which is largely undocumented and shared by teachers working in a given country's education system. The volume and cultural nature of this knowledge make it particularly challenging to share curricula and instructional methods between countries. Thus, approaches based on knowledge engineering—designing a software system by interviewing human experts to extract their knowledge and heuristics—are particularly promising for cross-cultural curriculum implementations. Reasoning Mind's Genie 2 system demonstrates the viability of such an approach, bringing elements of Russian mathematics education (known for its effectiveness) to the United States. Genie 2 has been adopted on a large scale, with around 67,000 United States students participating in the 2012–2013 school year. Previously published work (some of it in peer reviewed articles and some in non-peer-reviewed independent evaluations) has associated Genie 2 with high student and teacher acceptance, increases in test scores relative to “business as usual” conditions, high time on task, and a positive affective profile. Here, we describe for the first time the design, function, and use of the Genie 2 system. Based on this work, we suggest general principles—which collectively represent a proposed methodology—for the design of intelligent tutoring systems intended for cross-cultural transfer of curriculum and instructional methods.

Keywords Intelligent tutoring systems · Cross-cultural transfer · Mathematics education · Mathematics curriculum

Dedicated to the memory of Paul J. Sally, Jr

G. A. Khachatryan (✉) · A. V. Romashov · A. R. Khachatryan · S. J. Gaudino · J. M. Khachatryan ·
K. R. Guarian · N. V. Yufa
Reasoning Mind, Houston, TX, USA
e-mail: gak@reasoningmind.org

Introduction

Intelligent tutoring systems have operated through a wide range of different instructional mechanisms—i.e., different theories of how certain adaptive behaviors will lead to improved student learning. To name a few such instructional mechanisms, student models allow to appropriately allocate instructional resources (Koedinger & Corbett, 2006; Doignon & Falmagne, 1999), games engage students through choice and emotional involvement (Andersen et al., 2011; Ventura, Shute, & Kim, 2013), homework systems correct misconceptions quickly through immediate feedback (VanLehn et al., 2005; Razzaq et al., 2009; Pardos, Dailey, & Heffernan, 2011), and natural language processing interfaces promote active construction of knowledge (Graesser et al., 2003). Many systems combine two or more of these mechanisms; for example, the tutorial dialog agent AutoTutor includes both natural language processing and a student model (Graesser et al., 1999). Identifying and studying a broad variety of instructional mechanisms is important, since integrating multiple mechanisms may yield an additive effect on student learning.

In the case of school mathematics, one instructional mechanism that merits study is the underlying curriculum. Because mathematics is a sequential subject—with each new concept building on preceding ones—it is particularly sensitive to poor sequencing of topics or incomplete explanations. Countries that outperform the United States in international comparison studies have curricula that differ radically in content, sequencing, and mode of presentation from curricula used in the United States (Schmidt, Houang, & Cogan, 2002). Even the use of a more effective mathematics curriculum within a country—where curricula differ much less than they do across different countries (Stigler & Hiebert, 2009)—has had an effect size of 0.3 standard deviations on standardized tests (Whitehurst, 2009).

However, transplanting curricula across nations using conventional methods (i.e., translating and distributing printed materials such as textbooks) is a prohibitively difficult task (see, for example, Garelick, 2006). The style and mathematical approach of textbooks is intertwined with topic sequencings, lesson plans, and instructional methods, and for this reason the curriculum should properly be considered not just as the books, but rather—following G. Whitehurst's (2009) definition—as the sequence of instructional experiences intended to be delivered to students. Accordingly, the curriculum is interconnected with the professional preparation of teachers, the organization of schools, and even cultural aspects of the broader society. These systemic and cultural aspects of curriculum and instruction account for both the vast differences between countries and for the difficulties in international curriculum adoption (Stigler & Hiebert, 2009).

Viewing curriculum in this light, intelligent tutoring systems can be seen as a potential method for overcoming the obstacles to international curriculum adoption. By bringing instructional experiences informed by strong curriculum knowledge and pedagogical content knowledge directly to students, scaffolding teachers' learning of the new mathematical ideas and methods, and changing the classroom configuration to render existing cultural scripts inapplicable, one can attempt to implement in one country the curriculum methods developed in another.

This is the main goal of Reasoning Mind's Genie 2 intelligent tutoring system. Because curriculum is so intertwined with instructional methods, the approach requires

codifying the behavior of Russian mathematics teachers and then designing a learning environment—including an intelligent tutoring system as well as a set of practices expected of the United States teachers implementing the program—to simulate as closely as possible the combined instructional experiences these Russian teachers' students would receive.

While Genie 2 is not the first intelligent tutoring system to address curriculum, the topic has not received as much attention from researchers in the field as other questions. The work that has been done has mostly focused on studying specific topics or strategies. Salden et al. (2010) found that adding worked examples to Cognitive Tutors increased efficiency of learning. Rau et al. (2012) found that fluency and conceptual understanding were both beneficial in instruction of fractions with multiple representations, and Rau et al. (in press) extended this by arguing that conceptual instruction supported fluency training when it preceded it, while the opposite did not hold true. Adams et al. (2012) found that the use of erroneous examples in learning decimals in 6th grade improved performance on a delayed posttest. Rittle-Johnson & Koedinger (2009) found that alternating between lessons on concepts and procedures led to better learning of certain arithmetic operations than an approach where concepts were studied first, and procedures afterwards.

In contrast to these approaches, which use intelligent tutoring systems to study specific curriculum strategies, Genie 2 is an intelligent tutoring system that represents and delivers an existing curriculum in its entirety—i.e., as a system of interconnected pedagogical practices. In particular, Genie 2 illustrates how a curriculum tradition from one country (in this case, Russia) can be brought to students and teachers in another country. This is of practical significance, as Russian mathematics education is known to be particularly strong (as discussed in more detail below in “Russian Mathematics Instruction”).

Another approach to mathematics curriculum in the context of intelligent tutoring systems can be found in work done by Koedinger, Corbett, and colleagues at the Cognitive Tutor research group (Corbett, McLaughlin, & Scarpinato, 2000). On finding that school implementations of Cognitive Tutors were complicated by the difficulties of integrating with existing (paper and pencil) school curricula, the group developed core curriculum textbooks designed to work well alongside the Cognitive Tutor system. This approach shares with Genie 2 an integrative approach to curriculum, but differs in locating the introduction of new concepts primarily in the paper and pencil component of the program, whereas in Genie 2 the intelligent tutoring system is the primary vehicle for the introduction and discussion of new material.

The main theoretical contribution of the work presented here is to demonstrate that an appropriately designed intelligent tutoring system can be successfully used to implement aspects of the mathematics education tradition of one country in another, and to propose (in “General Principles for Cross-Cultural Curriculum Transfer using Intelligent Tutoring Systems”) a collection of design principles collectively representing a methodology for developing intelligent tutoring systems for such cross-cultural transfer. The main practical contribution of the work is to present a concrete such system capable of reflecting multiple aspects of Russian mathematics education, which (as discussed in “Russian Mathematics Instruction” below) is a particularly successful mathematical tradition. This system is notable in using detailed scripts (see “Instructional Content”) to capture and recreate the content-specific

strategies used by Russian mathematics teachers; this is one approach to meeting the challenge posed by Du Boulay and Luckin (2001) of developing intelligent tutoring systems that better capture the wide repertoire of teaching actions exhibited by expert teachers.

We begin with the “Methods” section, which sets the stage by describing the approach taken to designing and iteratively improving the intelligent tutoring system. Next, in “Russian Mathematics Instruction,” we introduce the main instructional principles that the system is meant to implement. The following three sections, “System Use and Behavior,” “Architecture and Knowledge Representation Structures,” and “Instructional Content,” describe the specific ways in which the system reflects those instructional principles. “Implementing Genie 2 in Schools” describes the use of the resulting system in the field, and “Outcomes” gives an overview of previously reported findings on the system’s effects and also presents new findings from a large-scale user survey. Next, “General Principles for Cross-Cultural Curriculum Transfer using Intelligent Tutoring Systems” outlines the general lessons learned from this work. In “Discussion” we consider the possible causes of the system’s positive effects on student outcomes, characterize the system as a whole incorporating both software and human components, and outline possible future directions. In the “Conclusion” section, we summarize the main contributions of this work. There is one appendix providing background on the history of the Russian mathematics education tradition, one giving some examples of scripts used to capture the Russian teachers’ practices, and one providing more detailed data from the user survey.

Methods

The approach taken in the present work was to compose a panel of several (approximately half a dozen) expert Russian mathematics teachers and to interview them individually and collectively in order to create a model of their approach. The members of this panel worked in a variety of schools, most (but not all) of them in Moscow. They were chosen through a combination of recommendations from Russian authorities on mathematics education and interviews, and included teachers from several of Russia’s best-regarded mathematics schools (such as Moscow School #56 and Moscow School #2) as well as faculty members of Moscow State Pedagogical University and Moscow Region State University. The members of this panel not only took part in interviews and discussions about Russian mathematics education and their own practices, but also participated in the making of software design decisions and wrote content for the resulting intelligent tutoring system.

The work was conducted as design research, with iterative cycles of development, piloting, and analysis (Collins, 1992). The system as discussed here went through eight such cycles, with the piloting phase of each one lasting a full academic year; the participating schools were all located in the United States, and varied in number from three (in the first year) to over three hundred (in the eighth year). As is typical for design research (Brown, 1992), the project sought both theoretical and practical outcomes: theoretically, to understand whether and how an intelligent tutoring system could be used for cross-cultural transfer of curriculum, and practically, to develop such

a system implementing the Russian mathematics curriculum in United States schools. The primary sources of information used were:

- (i) Interviews with and feedback from Russian teachers
- (ii) Feedback from United States teachers using the program
- (iii) Qualitative observations of classrooms using the program
- (iv) Quantitative data, including from system logs

This information was used iteratively, learning from each successive version of the system, modifying it, and then gathering information on the modified version. For example, Russian teachers typically begin each day with a warm-up activity, frequently a timed session over mental math. Early versions of the Genie 2 intelligent tutoring system modeled this by giving students a sequence of problems at the beginning of every day. After seeing the resulting system and visiting classrooms to observe students, the Russian teachers pointed out that this differed from their warm-up activities in several significant ways: the items weren't timed, the activity didn't feel to students like something distinct from the main flow of the lesson, and the feedback students received—detailed, step-by-step solutions—was too extensive. Accordingly, a new type of element—called “speed games”—was introduced and used in the warm-up activities; these were simple games that presented a sequence of mental math (or other fluency practice) items in an appealing visual form.

Feedback from United States teachers using the program was the most significant source of changes. Because of the large number of in-person site visits (see “Implementing Genie 2 in Schools”) by Reasoning Mind staff, strong relationships were built with teachers using the program. These teachers would provide constant feedback to support staff, who in turn would relay it to the development team. The most valuable teacher feedback was frequently provided not through written attitude surveys, but rather in person to support staff. Hundreds of changes were made based on teacher feedback; examples include the addition of a Spanish language glossary,¹ addition of sound (narration) to the materials, and the creation of new reports.

As the development team grew, the development process was increasingly aligned with user-centered design principles. In particular, significant changes to the student interface—such as the narration function mentioned earlier—were iteratively prototyped and usability tested before being implemented in the system. In the case of teacher interface functions, however, implementation and actual use—followed by cycles of classroom observations and iterative improvement—remained the most practical method.

Russian Mathematics Instruction

There is evidence that Russian students do well in mathematics, with Russian 4th grade students ranking 10th and Russian 8th grade students ranking 6th on the most recent (2011) Trends in International Mathematics and Science Study (Provasnik et al., 2012). In the International Mathematical Olympiad, the most important high school

¹ Many students in United States schools are first-generation immigrants from Spanish-speaking countries (Uro & Barrio, 2013).

mathematical problem solving contest, Russia's team has been one of the top five every year since 1999, with an overall record exceeded only by that of China (International Mathematical Olympiad, 2013). And in research mathematics, Russian mathematicians are heavily overrepresented: of the twenty-two mathematicians to have won the Fields medal² since 1990, seven—more than from any other country—have been of Russian origin. Teachers in Russia are also exceptionally strong, as Russian middle school mathematics teachers were found to have content knowledge second only to that of Taiwanese teachers (Tatto et al., 2009). Readers interested in a more detailed discussion of the history and international reception of Russian mathematics education can find this in the appendix, “Appendix 1: History and International Reception of Russian Mathematics Education.”

In the present work, interviews with the panel of expert Russian mathematics teachers demonstrated that there were numerous methods and approaches that they had in common, despite coming from different cities, having studied in different universities, and teaching in different schools. Furthermore, just as the methods used by the individual teachers were quite similar, the teachers confirmed that so were the approaches taken by virtually all of the mainstream mathematics textbooks available in Russia. For these reasons, it is impossible to draw a neat line between the various concrete Russian “curricula” represented by the mainstream textbooks, and instead one can speak of a single “Russian mathematics education tradition,” or even a general “Russian mathematics curriculum”—and indeed the teachers in the panel spoke in exactly such terms. We will follow their usage throughout this article, taking “Russian mathematics curriculum” to mean these general practices and knowledge shared by Russian teachers and transcending the particulars of any one textbook.

Next, we will discuss several of the most salient aspects of this tradition that emerged from the panel interviews, contrasting them with prevailing practices in the United States.

Guided Instruction The instruction provided by Russian teachers (especially when working with struggling or average students) is guided: students are carefully led through the material, rather than being encouraged to explore and arrive at their own conclusions.³ While Russian teachers do occasionally use techniques akin to discovery learning (in which children arrive at general conclusions through their own independent deductions), this is done sparingly, especially in the case of low-prior-knowledge students. This stands in distinction to practice in the United States, where pure discovery learning remains popular with many teachers and teacher educators (Mayer, 2004).

Fading In the course of work on a topic, teachers usually present material to students, then work examples on the board, then invite students to solve problems in a guided setting, and finally have students solve problems independently. This can be compared to “fading,” a practice for instruction through worked examples (Renkl, 2011). Actually, worked examples play a central role in Russian mathematics education: teachers

² This is the most famous award recognizing contributions to mathematics research.

³ This is in line with recommendations of cognitive load theory. See, for example, Mayer & Moreno (2010) or Kirschner, Sweller, & Clark (2006).

ask students daily to solve problems on the board in front of the class, and also require students to take notes with detailed problem solutions, subsequently to be studied while doing homework or in class. One of the findings of the present project was that this was less characteristic of classrooms in the United States: most teachers involved in the project did not have existing approaches to note-taking and the study of worked examples, instead having relied primarily on whole-class discussion of examples and subsequent use of worksheets. As such, while worked examples played a role in these teachers' practice, it was a less central one than in the practice of Russian teachers.

Motivating New Constructs New mathematical definitions and principles are motivated before being formally introduced. For example, before defining “factors,” a word problem could be given: “Among how many children can we divide 18 apples evenly?” Once the problem is solved, it becomes clear that the list of numbers (1, 2, 3, 6, 9, 18) is a mathematically meaningful one, and therefore deserves a name; it's at that point that the term “factors” is introduced and formally defined.

Analogously, general rules are introduced with motivation, too. For example, the teacher might motivate the fundamental property of fractions (that the numerator and denominator can be multiplied or divided by the same number without changing the fraction's value) by drawing an illustration showing that $1/3$ and $2/6$ are identical (e.g. using a unit segment subdivided into thirds and sixths), and then pointing out that we have merely increased the numerator and denominator by a factor of 2. This leads to the general statement of the property. As students mature, this kind of “proof by example” is replaced by actual mathematical proof. The transition happens mostly in grades 7–9, though elements of it appear earlier.

The goal of this method is for students to understand—or at least have an intuition for—the reasons for the mathematical notions and facts they encounter in the course. This forms an interesting counterpoint to the guided instruction approach taken by the Russian teachers: on one hand, students are explicitly led by the teacher through the material, but on the other hand, the teacher takes care to make sure that students do so with awareness of whence they are coming and where they are heading.

In United States classrooms, it is less common for teachers to motivate new constructs. For example, Stigler and Hiebert (2009) report that concepts in German and Japanese mathematics classrooms are far more frequently “developed” than in United States classrooms, where they are more likely to be introduced without such development (i.e. motivation).

Organizing Knowledge Around Mathematically Meaningful Structures Compared to teachers in top-performing countries, teachers in the United States often have weaker mathematics content knowledge (Ma, 1999; Tatto et al., 2009). This can lead to the organization of material around surface-structure patterns instead of around entities aligned with the deep structure of the subject. As an illustration of this, one school district the authors were familiar with went so far as to sequence 6th grade topics by operation: addition of whole numbers, decimals, and fractions, then subtraction of the three, then multiplication, and finally division. By contrast, Russian curricula are organized in a way that corresponds to the structures that experts in the subject (including mathematicians) generally consider mathematically important. For example,

each number system (whole numbers, rational numbers, real numbers) is studied as a whole, and the operations are grouped with addition and subtraction (which are inverse operations) together, and multiplication and division likewise (Milgram, 2005).

One significant example of this is the Russian treatment of common fractions and decimals. Rational numbers are known to be a difficult topic for students to learn (Behr & Post, 1992; Fuchs et al., 2013; National Mathematics Advisory Panel, 2008). In the course of interacting with United States teachers, it became apparent that many elementary school teachers introduced fractions and decimals independently, treating decimals as an extension of whole numbers. Indeed, from the standpoint of surface structure, this makes perfect sense: because operations with decimals are procedurally similar to those with whole numbers, a surface-structure grouping would place these two objects in one category, with common fractions in another. The problem with this approach is that it conceals the deeper mathematical structure: mathematically, (terminating) decimals are just a notation for a special case of common fractions—those whose denominators are a power of 10. By presenting and treating decimals as a special case of fractions—rather than an extension of natural numbers “to the right”—throughout the course, one can aim to develop in students an understanding of the rational number system as a single, unified whole. Early Russian curricula—such as the arithmetic textbook of A. P. Kiselev (see “Appendix 1: History and International Reception of Russian Mathematics Education”)—present fractions before decimals. However, there is a problem with this approach: operations with fractions are more difficult for students to master than those with decimals, and therefore studying fractions completely before studying decimals is less efficient. More recent curricula—for example, that of Vilenkin (Zhokhov, 2007)—introduce common fractions first, study the four operations with like fractions (and those that can be easily brought to a common denominator), then proceed to a complete study of decimals, and finally study common fractions (including the operations on general unlike fractions) in a complete fashion. This planning approach allows the curricula to present numbers as elements of number systems, while still sequencing things in a developmentally appropriate way.

Explicit Rules and Definitions Beginning in the 5th grade, Russian curricula organize the core declarative (or, as the Russian teachers call it, “theoretical”) knowledge of courses through explicit rules and definitions. This role of “rules” and “definitions” is analogous to the role of “theorems” (“lemmas,” “corollaries,” etc.) and “definitions,” respectively, in higher mathematics. For example, students learn to state the rule for adding like fractions: “To add like fractions, add the numerators and keep the denominator the same.” Being able to state these rules and definitions is seen as a pedagogical goal in itself: if a student can add fractions but cannot correctly state the rule, then the teacher will work with the student to teach him or her to do this. The rules and definitions are then used throughout the course; for example, if a student is unsure of whether something is or is not a ray, mixed number, greatest common factor, or other such mathematical object, then the teacher will refer the student to the corresponding definition and demonstrate how it can be used to make the determination. Likewise, if a student is unsure of how to follow a certain procedure, then the teacher will show how to read this procedure from the corresponding rule. Russian teachers see the

development of this skill—the ability to apply rules and definitions in specific cases—as an important independent goal of mathematics education.

In United States curricula and classrooms, by contrast, there is no such analogous role of rules or definitions. While terms are defined and procedures described, less attention is often given to the precision of the wording (Givental, n.d.; Milgram, 2005; Wu, 2011). This is corroborated by findings from the present project, in which it was seen that teachers were not familiar with the use of such a “theoretical backbone” of a mathematics course.

Mathematical Precision Statements made in a mathematics course—and especially rules and definitions—are formulated in a mathematically precise way. (Mathematicians call this “mathematical rigor.”) For example, saying that “a factor of a number n is a number that divides n evenly” is not precise, since this would lead us to conclude that 1.5 is a factor of 3. Because this definition is meant to apply only to natural numbers (meaning positive integers), and because only natural numbers can qualify as factors, the definition can be fixed by saying, “A factor of a natural number n is a natural number that divides n evenly.”

Russian teachers and curriculum authors take substantial care to ensure that all of their statements are precise and use correct terminology. The teachers also correct students when they make mathematically imprecise statements.

Explanatory Feedback While some activities only provide corrective feedback (most notably, exercises meant to develop automaticity over already acquired skills), the vast majority of feedback given by Russian teachers is explanatory (cf. Hattie & Gan, 2011). In particular, explanations refer back to underlying definitions and rules whenever possible. For example, when asked to give the smallest multiple of 9, some (grade 6) students will answer that it is 18. In this case, a Russian teacher might correct this by asking, “What’s the definition of a multiple?” The student (or another student in the class, if this one doesn’t remember it) would answer, “A multiple of a natural number n is a natural number that can be divided evenly by n .” Then, the teacher would ask, “Is 18 the smallest number that can be divided evenly by 9?” The student would be expected to then realize that the correct answer to the original question is 9, not 18.

Automaticity Through Spaced Practice Russian teachers view concepts and skills as mutually reinforcing. In fact, the word they use which is closest to the English term “knowledge component” is “ZUN,” which is an acronym for “znaniye, umeniye, navyk,” translating roughly as “knowledge, skill, automatic skill.” In the Russian teachers’ way of thinking, each knowledge component contains some combination of these three things; in knowledge components that include an “automatic skill” part, acquiring automaticity is seen as essential to mastery of the ZUN.

The main vehicle for automaticity is spaced practice. Specifically, the first several minutes of each class are usually spent doing timed mental math or calculation problems over already studied topics. This serves a dual purpose: practicing skills, and warming students up for the new material to be presented that day.

This view—namely, that the development of automatic skills reinforces the formation of new concepts—is in line with findings in cognitive psychology. For example, in cognitive load theory, the automation of prerequisite skills frees working memory for the construction of new concepts (Kalyuga, 2010). However, it is far from universal in United States schooling, where there has been a debate between a “skills” camp and a “concepts” camp (Wu, 1999).

Breaking Concepts into Constituent Components Sometimes, a certain concept is difficult to learn, but can be broken up into components that can be studied separately and then “reassembled.” Russian curricula frequently resort to this strategy.

Perhaps the most salient example in the early grades is the approach to solving equations. To solve equations, students must—among other things—(i) be familiar with unknowns, and as comfortable operating with them as with numbers, (ii) be fluent in applying the properties of the operations to transform expressions, and (iii) understand the equals sign as a relational symbol.⁴ Russian curricula resolve issue (i) by introducing unknowns as early as 2nd grade, asking students to solve simple equations and to evaluate algebraic expressions at various values of the unknowns. As for issue (iii), this is addressed both through the constant recurrence of equalities with operations on both sides (such as $1+2=2+1$) and through the teaching of explicit definitions of the equals sign as a relational symbol: “An equality is a mathematical statement that two expressions are equal.” But the most interesting approach is the one to issue (ii). Here, the difficulty is that it is hard to justify manipulating expressions without any unknowns; Russian programs (such as Vilenkin) get around this by giving “convenient calculation” problems: students evaluate expressions like $98+57+2$ or $13\times 57+87\times 57$, and justify their answers by naming the properties they have used.

In the language of cognitive load theory, Russian curricula manage intrinsic cognitive load through pre-teaching (cf. Mayer & Moreno, 2010).

Placing the Concrete Experience of Expert Practitioners Above the Application of Any General Principles This last principle of Russian mathematics teachers is perhaps the most important, and serves as a kind of “anti-principle.” The Russian mathematics teachers interviewed all believed that mathematics teaching was an experimental craft, and that lesson planning and the selection of approaches to explaining certain concepts were very sensitive to the particulars of the topic, curriculum, and individual students making up the class. Therefore, while they relied on certain principles (including the ones listed above) for general guidance, the overriding principle was to make pedagogical decisions based on the accumulated experience of the community of Russian mathematics teachers, which experience is seen by Russian teachers as almost entirely content-specific.

This is in contrast to approaches taken in the United States, where content-specific knowledge is less universally valued (Ma, 1999; Shulman, 1987; Hill, Rowan, & Ball

⁴ Unless taught otherwise, students in the early grades often see the equals sign as an operational symbol only – i.e., a command to perform the operation indicated to the sign’s left (McNeil et al., 2006; Baroody & Ginsburg, 1983).

2005). Furthermore, the public and political dialog surrounding mathematics instruction in the United States has frequently centered around debates between competing camps of “isms,” as described long ago by Dewey (1938) and evidenced more recently by the so-called “math wars” (Ravitch, 2001; Schoenfeld, 2004).

System Use and Behavior

Genie 2 is intended for students in grades 2–6. It is used in classrooms, with each student sitting at their own computer and the teacher walking around the room to monitor the class and work with individual students or small groups (Fig. 1). Students work online at their own pace, so on a given day different students study different topics. Nevertheless, students sometimes discuss their assignments with one another. Ocumpaugh et al. (2013) found that on average, students using Genie 2 spent 7 % of their time in on-task conversations.

The student home screen (Fig. 2) is a map with a collection of buildings, each representing a different mode. Mode availability is determined by the system, and teachers can further restrict access to certain modes. Students then select among the remaining modes. The most significant modes are:

- *Guided Study*. This is the main learning mode. Here, students study their assigned course, which consists of a sequence of “objectives.” The content within each objective includes explanatory (called “Theory”) material, basic problems (called “level A” problems), and—for students who do well on the basic level—more advanced problems (at difficulty levels B and C). Students spend the overriding majority of their time in this mode.
- *Homework*. The system selects homework problems based on the student’s place in the curriculum and performance on prior topics; these problems are printed out by the teacher and done on paper at home by the student, who can enter them into the system the next day for automatic grading and feedback.
- *Online Tutoring*. This mode allows students to receive online tutorials from human tutors. (In practice, this mode has been used only by a small—albeit increasing—fraction of students studying in Genie 2, since few schools are willing to pay for online tutoring.)
- *MathRace*. This is a two-player game in which students race to answer arithmetic questions.



Fig. 1 A class using Genie 2



Fig. 2 The Genie 2 student home screen

- *Office*. Here, teachers can assign individual topics from Guided Study, standardized test practice problems, or entire practice tests.
- *Review Mode*. In this mode, students are given a sequence of problems from all of the objectives they have already covered in Guided Study. The problems are chosen to emphasize topics on which students have struggled. If students do well in the mode, the difficulty level of problems increases; there are six different levels, each of which has a different mix of problem difficulty, from 100 % level A (level 1) to 100 % level C (level 6). When students are in Review Mode, their current level in the mode is displayed at the top of the screen.
- *Wall of Mastery*. This mode (Fig. 3) was added as a systematic way for students to learn to solve level B and C problems on each objective presented in Guided Study. Once an objective is covered in Guided Study, the level A cell in the corresponding row in the Wall of Mastery becomes enabled. If the student selects a cell, a sequence of problems from the corresponding difficulty level and objective are given; if the student completes the problems with high accuracy, the cell is marked completed and the cell of the next difficulty level is enabled.⁵

In Guided Study, students normally progress linearly through the sequence of objectives. However, it is possible (and in practice frequent) for a student to be diagnosed to be lacking in prior knowledge and sent back to earlier objectives.

Each Guided Study objective contains a number of modules, including Warm-up (simple practice problems or games, usually involving mental math), Theory (explanatory material), and Level A Problems. Each objective also has its own Level B Problems and Level C Problems modules, a Review module (which includes several problems over previous objectives), and a Notes Test module to verify that students took good (pencil and paper) notes during the Theory. Poor performance on certain

⁵ The Wall of Mastery can be regarded as a rudimentary open learner model. Bull (2004) has found that students respond positively to the idea of open learner models, and Kay et al. (1997) argue that such an approach is important for self-regulation. Because learner control is often more effective in situations when learners have more prior knowledge (Clark & Mayer, 2011), it is likely more appropriate to use such an interface for advanced problem solving over previously studied topics—as in the Wall of Mastery—than in a basic instructional mode such as Guided Study.



Fig. 3 The Wall of Mastery mode

modules can lead to being sent back to previous modules of the objective. The Level B Problems and Level C Problems modules are enabled if students are on schedule in Guided Study and perform well in the Level A Problems module. To let students chart their progress, a map (Fig. 4.) is displayed to students between modules in Guided Study. This map includes a site for each module along with pathways showing when students can be returned to prior points.

Teachers using the system frequently requested sound support for weaker readers, and this function was subsequently introduced into the system: now, students can click a speaker button to hear the most recently appeared phrase read out. Natural voices are used for Theory material, and synthetic ones for problems.

Genie 2 has a number of game-like elements. In addition to points (earned for correctly solving problems), the system includes a “streak” (tracking the number of problems solved correctly in a row) and a virtual prizes system allowing students to redeem their points for e-books, animations, and decorations for a virtual room (called “My Place”). These game-like features are analogous to several of the features of iSTART-ME, the “motivationally enhanced” version of iSTART, an intelligent tutoring system for reading (Jackson, Boonthum, & McNamara, 2009).

Another game-like element of Genie 2 is the Genie (Fig. 5), an agent who congratulates students on correct problem answers and encourages them throughout their work



Fig. 4 The topography of an “objective”



Fig. 5 The Genie: pedagogical agent, Genie 2 namesake

in the system. The Genie's personality is built in throughout the system: students can visit the Genie's house, buy Genie-themed virtual prizes, and send emails to the Genie; a team of part-time writers compose responses to these emails. Pedagogical agents in other online systems—such as Design-a-Plant—have been shown to increase both learning and enjoyment for students using these systems (Lester et al., 1997).

In addition to a student interface, Genie 2 has an interface for classroom teachers. This interface has several reports, including an Objective Spreadsheet Report that displays a grid with students as rows, objectives as columns, and problem accuracy (or other useful measures) in the cells, an Activity Logs report that allows teachers to view problems incorrectly solved by a student and generate similar problems for interventions, a dashboard with summary data on each student and the ability to assign work and adjust settings, and a Notifications Report that gives a live feed of information on students' successes and failures in the system.

Architecture and Knowledge Representation Structures

The behavior of Genie 2 was determined by interviewing the panel of expert Russian teachers to understand their practices, designing corresponding system algorithms, and then refining these algorithms through collaborative discussions with the teacher panel. The most important algorithm for the system's behavior is called the *decision system*. The decision system operates in concert with a student model, which holds a wide range of information about the student's knowledge and other instructionally relevant attributes. The decision system regulates access to different learning modes, updates the student model, and selects content to be delivered to the student (Fig. 6.). The use of a separate student model and pedagogical model (in this case, the decision system) is analogous to approaches taken in some tutorial systems; for example, Ms. Lindquist is a tutorial system that combines a student model with tutorial planning behavior developed based on observations of an expert tutor (Heffernan & Koedinger, 2002). As another example, in the domain of cardiovascular physiology, the intelligent tutor CIRCSIM was designed based on typed transcripts of actual tutoring sessions and explicitly intended to simulate expert tutors (Evens et al., 1997).

Instructional content includes specific problems, explanatory animations, and instructional games, as well as the modules assembling these atomic objects into the clusters in which they can be assigned. By contrast, the instructional data structures

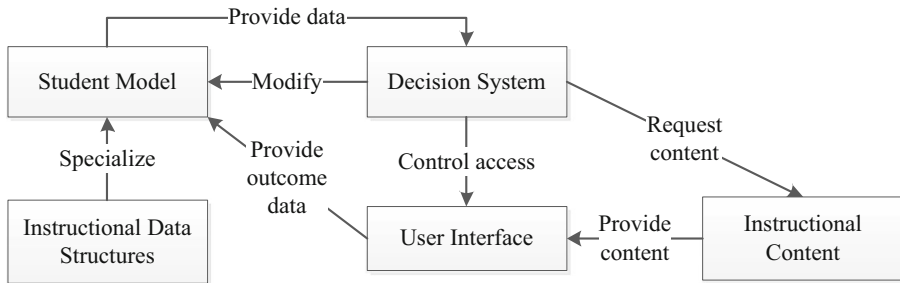


Fig. 6 Main components of the Genie 2 architecture

(which include things like calendars and course sequencings) serve to arrange information about the course and individual students in a meaningful fashion.

Information in Genie 2's instructional data structures, student model, and instructional content is organized through object-attribute-value triplets. The primary instructional data structures are the "objectives" introduced in the preceding section. Objectives serve a dual purpose: they are both knowledge components (each of which is relatively large, such as addition & subtraction of like fractions, or the perimeter of a rectangle) and units of study, in that content modules are organized by objective. Objectives are equipped with *prerequisites*, thus forming a directed cycle-free graph. This is analogous to the data structures used in knowledge spaces (Doignon & Falmagne, 1999).

In systems built on knowledge spaces, the prerequisite relations are used both for diagnostics, and to allow learners to take different paths through the material. In Genie 2, by contrast, only the former holds: while students can be sent back by the decision system to previous objectives, the regular course of study happens according to a preset *sequencing*, which is a linear ordering of objectives respecting the prerequisite relation. This design decision was made to better model the instructional strategies of Russian mathematics teachers: the teachers aim to build multiple connections between mathematical topics, even when these topics aren't tied through the prerequisite relation. For example, in studying the addition and subtraction of decimals, Russian curricula will incorporate problems on finding the perimeter of a rectangle with decimal sides; this means that, even though the perimeter of a rectangle is not a mathematical prerequisite for the addition of decimals (and in particular it would be incorrect to suspect a child of not knowing perimeter based on a failure to master decimal addition), it is nevertheless necessary to ensure that students have studied perimeter before they are taught this topic. Organizing objectives into a sequencing is one way to resolve this problem. This need for a sequencing object illustrates how differences in instructional approaches can have implications for even the most basic data structures.

Another important instructional data structure is the *calendar*. This object gives the ability to store the schedule of planned online sessions.

These instructional structures—objectives, sequencings, and calendars—are then specialized to corresponding objects in the student model. Thus, each student is assigned a specific calendar and sequencing, allowing student knowledge to be tracked and stored by objective; in addition to an overall status ("not yet met," "in progress," "met," or "diagnosed"), each objective has a numerical knowledge level indicating the

degree of proficiency. The calendar and the statuses of objectives together can be used to calculate how far behind or ahead of schedule a student is, since each objective has an attribute giving the expected number of hours required to study it.

In addition to these conceptually meaningful objects (coming from the instructional data structures), the student model module contains all of the raw data generated by students in the course of their study in the system: problems given, answers provided, time spent online, and so on. While the conceptual objects in the student model are more frequently relevant to instructional decisions, the raw data is also sometimes of value: for example, in giving a student a problem, it can be useful to select a version of the problem (i.e., a “dataset”) which the student has not previously seen.

Much of the content-general logic of Genie 2 is contained in the decision system, which has the goal of capturing the reasoning process used by the Russian teachers. The decision system:

- (i) Determines which instructional modes should be available to the student at each point in time.
- (ii) Analyzes raw student data and makes corresponding changes to the student model (including updating the status of objectives).
- (iii) In each learning mode, analyzes the student model and selects corresponding instructional materials.

The logic of the decision system was developed based on interviews with Russian teachers. While the initial plan was to implement all of the logic through explicit rules, it became clear in the course of interviewing the teachers that some of their reasoning was more procedural in nature, and as such lent itself better to implementation through conventional (procedural) algorithms.⁶ As a result, the decision system was built to contain both a procedural component and a collection of explicit production rules.

The production rules were written in pseudocode; a compiler was developed to interpret the rules, producing a java program that was then included in the system build and executed.⁷ As an example, one rule in the pseudocode looks like this:

```
IF (student_position) is {on_schedule, ahead_of_schedule}
AND (student_level is {at_grade_level, above_grade_level})
THEN how_many_problems_can_be_assigned_to_student = many;
ELSE how_many_problems_can_be_assigned_to_student = normal;
```

⁶ For example, the process of studying a chosen topic is viewed by the teachers in a very procedural way: there are certain modules that must be studied in a certain order, with specified branches depending on local outcomes. By contrast, determining the learning level of an objective or conducting diagnostics and remediation are tasks which the teachers approached in a substantially more rule-oriented manner.

⁷ A practical aside: initial builds of Genie 2 directly ran the java code compiled from the rules, and this created performance problems as the number of students grew from year to year. Consequently, the compiled code was manually refactored to preserve the functionality while optimizing performance.

The rules are grouped into rule sets, with each set corresponding to a particular subtask. The decision system works in a data-driven fashion: i.e., at each step in a student's work in Genie 2, either a specific rule set or the entire collection of rule sets is executed afresh, reviewing the student model and taking action based upon it. This aspect of the architecture is particularly important because Genie 2 contains a number of different learning modes, and a data-driven approach allows students to move at any point in time from one mode to another and resume their work there.

The two major methods for reasoning with production rules are forward chaining and backward chaining. In the former, one reasons from data to conclusions, and in the latter, from goals to data (starting with desired consequents and then checking if there is data to support them). The development team initially hypothesized that backward chaining would be a more natural method of reasoning for modeling the teachers, since teachers were expected to reason from pedagogical goals; surprisingly, the panel interviews showed that the teachers predominantly reasoned from student performance to the student model, and thence to instructional actions. (For example, a teacher might draw certain conclusions about a student's knowledge based on his or her performance, and then use this knowledge to adjust previously planned subsequent learning tasks to better suit the student's needs.) Accordingly, the rule engine was implemented with forward chaining.

This decision illustrates how specific aspects of a certain pedagogical tradition can have implications for the architecture of intelligent tutoring systems designed to model them. In this case, the approach to pedagogical reasoning taken by Russian teachers influenced not only the *specific rules* needed, but also the underlying *inference model*. For example, it was found in the course of working with teachers implementing Genie 2 that many mathematics teachers in the United States view instruction through the prism of state standards (cf. Hamilton et al., 2007), and teach these standards in a certain (sometimes mathematically arbitrary) order, spending more time on—and returning multiple times to—those which students master at a lower level. For example, the most common request received from participating teachers was to “allow reordering of the objectives to match the district sequencing,” since this would allow teachers to use the district-administered “benchmark” (high-stakes test simulation) tests to identify standards requiring additional attention. This is in contrast to the Russian approach, which seeks to tell a single, cohesive mathematical story over the course of the year, with each topic building on preceding ones. It is possible that this emphasis on curriculum coherence constrains Russian teachers' inference patterns to forward chaining. This hypothesis—that teachers working in different curriculum traditions rely on different inference patterns to varying extents—has not, to the authors' knowledge, been researched yet, and would be interesting to investigate.

The use of explicit production rules to model teacher reasoning is not new: the approach dates back to at least the 1980s (Clancey, 1987), though it subsequently became less popular. Significant contemporary intelligent tutoring systems using explicit rules include AutoTutor, a tutorial dialog agent that uses a collection of production rules to respond adaptively to student input (Graesser et al., 1999).

The most salient component of the procedural part of the decision system algorithm is the workflow inside each objective (Fig. 7). This algorithm reflects many aspects of

Russian teachers’ practice: beginning each session with a warm-up activity, fading (in this case, beginning with Theory and then moving to Problem modules only once this has been successfully passed), ensuring that students take written notes, and reserving more difficult problems over a given topic for those students who have successfully learned to solve simpler ones.

Instructional Content

The Genie 2 system has core 5th and 6th grade curricula, each of which takes about 120 h for students to complete. In addition, there are smaller curricula for grades 2–4, designed to take 70 h each to complete; these curricula are usually used as ancillary programs, but sometimes—when supplemented with paper materials—as the core.

Once the sequence of objectives for a course (to be delivered in Guided Study) has been mapped out by the team of expert Russian teachers, each segment of related objectives is assigned to a teacher. For each objective, the teacher writes a detailed script describing how the teacher would present that objective; this script, while explicitly written for the online lesson, reflects the decision-making process the Russian teacher uses in classroom instruction. These scripts include both material (such as text and illustrations) meant to be explicitly presented to students, and instructions (usually

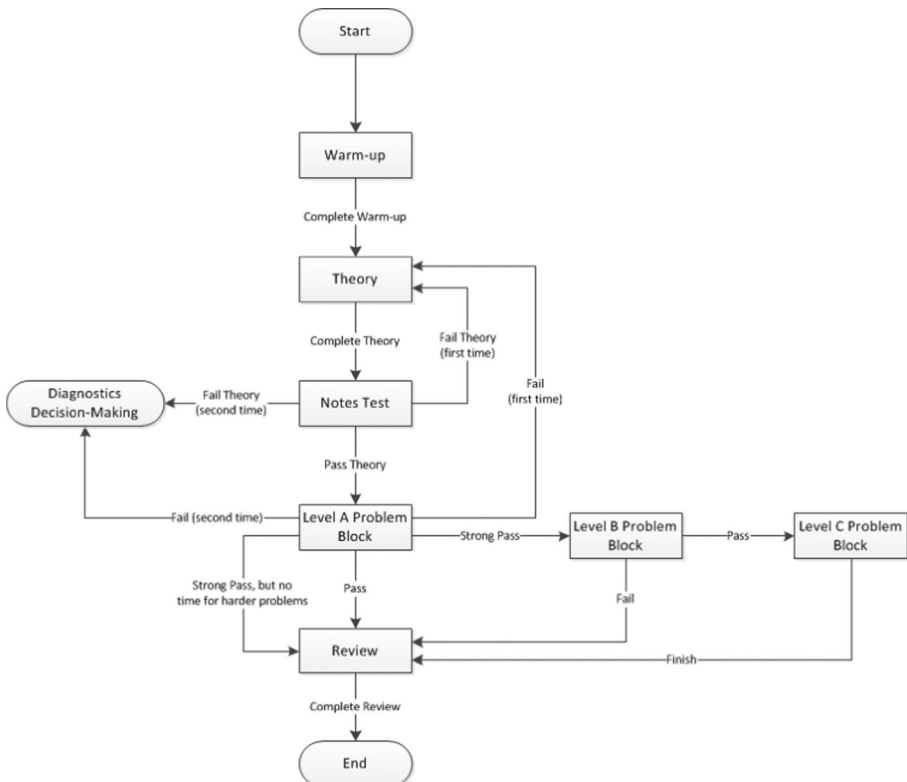


Fig. 7 The main flow inside an objective in Guided Study

given in brackets) for which questions to ask students and how to respond to different answers.

Initial versions of Theory (i.e., explanatory) content scripts were written by the Russian teachers in a style akin to textbook chapters. These early materials mostly consisted of one-way presentations of screens and animations with little asked of the student besides clicking a “play” button to move on. In fact, this was a failure of modeling, as Russian classroom instruction is very interactive. As a consequence of this failure, most students did not make an effort to read Theory material presented to them. So, it was decided that the system should model teachers better in this regard, interactively assessing every element that students were expected to attend to. In practice, this meant that students were asked questions on every screen, or roughly every half minute. These questions in the Theory material did not grant points, but still had the desired effect on students.⁸

The scripts written for these more interactive materials contained frequent questions, with different responses triggering a different reaction. For example, a Russian teacher might make a certain point, ask a question, and then react differently depending on a student’s response—moving on if the question was answered correctly, explaining the point again if it was answered incorrectly, or commenting on a specific misconception if the student’s answer suggested it was present. There is no single consistent way that teachers respond to students, as their behavior depends on the mathematical content. Figure 8 shows an example (translated into English) of an excerpt from such a script.

These scripts capture a number of aspects of Russian teachers’ practices, including the mathematical flow of their explanations of concepts, the collection of problems assigned to students, the manner in which students are questioned, the approach to developing metacognitive skills (in the case of Fig. 8, checking one’s answer), the specific wordings and visual representations used, and the ways in which teachers respond to different student answers. However, the level of detail of these scripts varies considerably from teacher to teacher, especially in the degree of interactivity. Branches in behavior are often indicated, but without great detail; an example of this can be found in Fig. 8, where the teacher means to treat differently the cases of a correct and an incorrect answer.

After a teacher completes a script, it is translated into English and sent to a curriculum developer. The curriculum developer substantially revises the script, making it culturally appropriate to an American audience, adding animated and game-like elements, introducing additional interactivity (especially in places where the teacher would likely have interacted with students, but omitted this from the script), and structuring the document to ensure that it is unambiguous for the artists and programmers developing the content. The curriculum developer aims to elaborate the script while fully preserving the mathematical and pedagogical structure; a lead teacher (who reads English) reviews the revised script to make sure that this has been accomplished. Some examples of completed scripts can be found in an appendix, “Appendix 2: Two Excerpts from Lesson Scripts.”

⁸ This observed phenomenon is consistent with M. Chi’s interactive-constructive-active-passive (ICAP) framework (Chi, 2009), and suggests that—in contrast to answering closed-ended questions – simple clicking on a screen should be classified as a “passive” activity, since it does not present real evidence of attending to the material.

Let's solve the equation
 $12 + x = 83$
 This is an equation with [list-box an unknown summand / an unknown minuend].
[The correct answer is "unknown summand." If an incorrect answer is given, then show the correct one.]
 In order to find an unknown summand, it is necessary to subtract the known summand from the sum.
 Let's select the right expression for solving this equation.
[Next->] In order to find x, we have: [list-box $x = 83 + 12$ / $x = 83 - 12$ / $x = 83 - 83$]
 Let's calculate the value of x.
 X = [entry field]
 Let's check whether we solved the equation correctly. We'll substitute the number <number entered by the student> into the equation in place of x.
 ?
 $12 + 71 = 83$
 We obtained a <true/false> equality.
 We obtained a true [or, in the case of an incorrectly entered number, "false"] equality. This means that 71 is a solution of the equation. [In the case that the student entered an incorrect number, give that number instead and say "is not a solution of the equation."]
[If the answer 71 was entered by the child incorrectly, then the text:]
 Let's calculate 83-12
 x=[uINT]
[The correct answer is 71. If the answer is incorrect, then show the correct answer and carry out checking for the correct answer.]

Fig. 8 An example of an objective script written by Russian teachers

The problems are implemented as instances of a general “problem” object; each problem includes a statement, attributes describing difficulty level and prerequisite knowledge needed, a full step-by-step solution (often with animated illustrations or explanations), and rules for checking answers (since input formats vary, including free numerical response, formulas, and geometric construction). Everything besides problems—including explanations, illustrations, and scaffolded exercises—is implemented as a sequence of Flash animations in the Theory material. (Two examples of Theory screens from the 4th grade curriculum are shown in Fig. 9). Other elements that can be

Fig. 9 Examples of Theory content from the Genie 2 Guided Study mode

present in the script include visual illustrations, animations, and games. Because of the large quantity of branching, the logic of each Theory animation can be relatively complex, and an animation taking just a few minutes for the student to view could have several thousand lines of code.

All told, the content in Genie 2 simulates the work of Russian teachers with a certain degree of accuracy. In addition to the “locally adaptive” behavior of the Theory (which captures many aspects of the dialog Russian teachers have with students), Genie 2 inherits from the Russian teachers their approaches to explaining mathematical concepts, the coherence of the curriculum sequencing, and the selection of mathematical problems. Many of the problems are very challenging, especially at the B and C difficulty levels. For example, Fig. 10 shows a level C problem from the 2nd grade curriculum.

Because so much of the instructional logic in Genie 2 resides in the content scripts, the resulting system is very sensitive to instructional differences between grades, topics, and even different components of each topic. Other systems that have captured teachers’ content-specific knowledge include Ms. Lindquist (Heffernan & Koedinger, 2002) as well as REDEEM (Major, Ainsworth, & Woods, 1997), a system that gave teachers a simple interface to edit instructional content, building in question prompts and other content-specific logic. The approach taken in Genie 2 differs from REDEEM in taking the content-specific logic from a small group of expert teachers and in imposing far fewer limitations on specific forms of logic possible in the content. Thus, the Genie 2 approach imposes greater costs for developing content, but also allows a wider variety of instructional strategies to be employed.

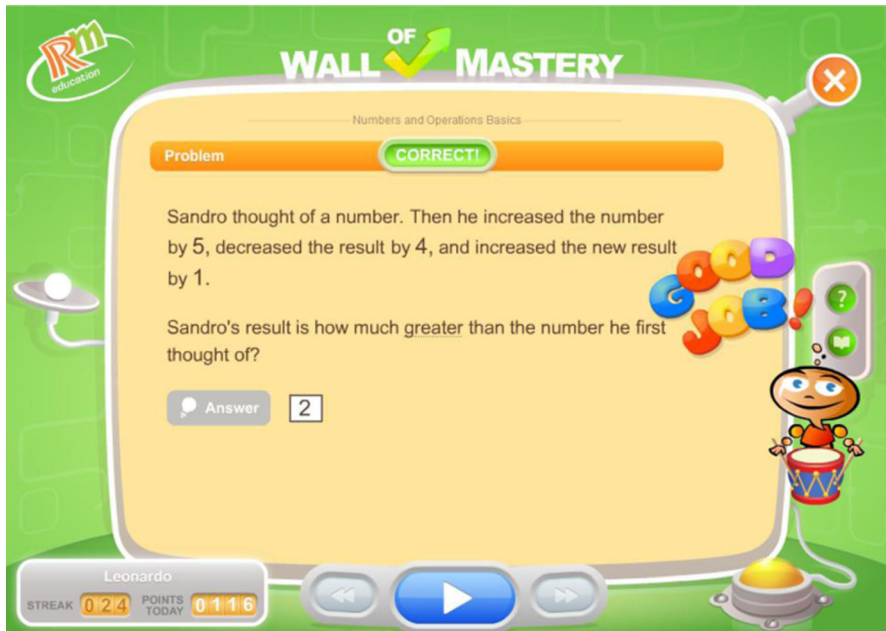


Fig. 10 A response is shown to a correctly solved 2nd grade level C problem

Implementing Genie 2 in Schools

Genie 2 was first used in the 2005–2006 school year. In this first year of use, system implementations in schools were of uneven and generally low quality (Weber, 2006). Several factors were identified as likely contributing to this. First of all, the software was complex, with many different modes and functions in the student as well as teacher interface. Second, teachers were unaccustomed to the new classroom configuration when using Reasoning Mind, and consequently did not always immediately understand their role in the classroom; in particular, many teachers did not understand that interventions with individual students or small groups were needed. Third, the strategies required to motivate students and manage the classroom were different from those in a traditional classroom, and teachers did not know these strategies and did not have the necessary context to invent them. And fourth, a lack of knowledge of Genie 2 curriculum content made teachers uncomfortable doing interventions and limited the quality of these interventions.

To remedy this, teacher professional development and support were made a required part of the program beginning with the 2006–2007 school year. The first year's courses were developed by Reasoning Mind staff, with each subsequent year incorporating changes based on feedback from teachers and school support staff. Three years after the first courses were offered, a separate department was created to develop and administer professional development. Several staff members of this department are former Reasoning Mind teachers.

As of the 2013–2014 school year, the professional development includes either a 10-h online course or a 2-day in-person course taken in the summer before the first year of implementation, along with six workshops (done in groups of several teachers) over classroom strategies and curriculum content throughout the year. Additionally, teachers study the curriculum to understand the sequence, concepts, and approaches to explaining mathematical ideas taken in the program. Teachers take an exam after the summer classes as well as three curriculum knowledge assessments throughout the year.

The emphasis on mathematical content knowledge is necessary for several reasons. First, it is important to motivate the use of the system for teachers, who often require additional content knowledge in order to understand the ways in which the curriculum is stronger than what their school district was using previously. Second, teachers must be comfortable with the curriculum material in order to be able to conduct effective one-on-one or small group interventions with students. And finally, curriculum and content knowledge are worthwhile in themselves, since mathematics content knowledge is important for effective teaching, and teachers in many countries—including the United States—often lack a strong preparation in this area (Ma, 1999; Hill, Rowan, & Ball, 2005).

As to the support, each teacher is assigned an implementation coordinator. Implementation coordinators are trained by Reasoning Mind, and thus do not necessarily need prior experience in education; some implementation coordinators join directly following college or graduate study, though an increasing number join with several years of experience as classroom teachers—a background which is not essential but which has been found to be helpful. A teacher's assigned implementation coordinator answers the teacher's questions and helps the teacher select suitable classroom strategies. Implementation coordinators also visit teachers' classes to observe teachers and give them feedback on their use of the program. On average, teachers receive six visits throughout the year: three observations and three discussion meetings.

Implementation coordinators conduct observations using a detailed protocol, rating teachers' performance in a number of categories. Teachers are marked as "Not Yet Established," "Established," "Proficient," or "Advanced" in each category. The categories are:

- *Data-driven decisions.* Using the system's reports to make instructional decisions.
- *Lesson planning.* Planning out lesson activities beforehand, including student interventions.
- *Instructional methods.* Conducting interventions; to achieve higher levels, the interventions must be varied and must include strong as well as struggling students.
- *Learning modes.* Spending a large fraction of scheduled time online; regularly using the Wall of Mastery and Review Mode.
- *Teacher engagement.* Being directly engaged with students during class time.
- *Procedures.* Having smooth class procedures, ensuring productive use of class time and good student behavior.
- *Incentive systems.* Setting clear individual student and class goals and rewards.
- *Notebooks.* Ensuring students take notes and show their work in well-organized notebooks.
- *Independent learning.* Ensuring students know Reasoning Mind's independent learning strategies (including that when a student solves a problem incorrectly, the student should carefully read and understand the problem's solution).
- *Student engagement.* Keeping students on task.

Once a teacher achieves a certain level in every category, the teacher is certified at that level. For example, a teacher who is "Proficient" in most categories and "Advanced" in the remainder would be certified at the end of the year as "Proficient."

The classrooms of teachers in different categories appear quite different. While time on task (as discussed below) is generally very high, it is particularly high in classes of "Advanced" and "Proficient" teachers. Teachers in these categories have strong classroom management, coupled with incentive systems to motivate students.

Of course, classrooms implementing the program vary substantially in their approach, in no small part owing to the relatively large scale of the program's use. Approximately 24,000 students used the program in the 2010–2011 school year, 47,000 in 2011–2012, and 67,000 in 2012–2013. In 2011–2012, the program served 379 schools in 56 districts in 8 states; students spent 2,051,542 h online, during which time they solved (correctly or incorrectly) 44,802,848 problems; 1,282 teachers participated in the professional development, and thirty-four full-time implementation coordinators supported teachers with in-person visits throughout the year.

Outcomes

In this section, we survey previously published results concerning different aspects of the system's effectiveness. This includes (non-peer-reviewed) reports written by external independent evaluators as well as a peer-reviewed study conducted by R. Baker's group on behavior and affect in Genie 2 (Ocumpaugh et al., 2013). We also present previously unpublished results of a survey of students, teachers, and principals using the system.

Genie 1 Learning Outcomes A prototype of Genie 2, this system was implemented in 2003 at an urban middle school in Texas. The curriculum focused on ratios and proportions. The implementation was in one classroom (30 students), with the system designers present for every session. Students used the system for 90 min every other day in addition to their regular mathematics class. The study was a randomized controlled trial, with 26 students in the control group. The program's impact was evaluated by Weber (2003). The study found an effect size between the experimental and control group of 1.66 standard deviations on an experimenter-designed test (which covered only ratios, rates, and proportions) and 0.79 on the Texas standardized test, the TAKS (which covered all 7th grade topics required in Texas).

Genie 2 Learning Outcomes One challenge in assessing a United States program based on the Russian curriculum is that most available instruments are much more proximal to United States curricula than to Russian programs.⁹ For example, a study of three implementations in the 2006–2007 school year (Waxman & Houston, 2008) compared students using the program to students studying the “business as usual” curriculum at their school. The study—which consisted of three concurrent yearlong randomized controlled trials at three separate schools, with several classes of students participating in each school—found that between-group effect sizes on an instrument developed by the experimenters were in the range of 0.4–0.8 standard deviations, while differences between experimental and control groups on the Texas high-stakes test were not statistically significant.

The experimenter-developed test consisted of items written by Russian teachers; these items were free-response, and reflected the traditional Russian emphasis on core arithmetic, algebraic concepts, and succinct language in word problems.

After practice standardized test problems were added to the program to facilitate transfer of learning to this format, another study by the same authors found a small but statistically significant effect of the program on scores on the Texas Assessment of Knowledge and Skills test (Waxman & Houston, 2012), with a between-group effect size of $d=0.14$. This study was a yearlong quasi-experimental design with 637 students in the treatment group and 777 students in the comparison group. As in the preceding study, control group students received “business as usual” mathematics instruction.

Affect and Behavior Outcomes Another significant outcome of the system's use is very high time on task and positive affective profile of students. Ocumpaugh et al. (2013) used the Baker-Rodrigo Observation Method Protocol (BROMP) to observe students using the program; in observations of three Texas schools—two of which served predominantly economically disadvantaged, minority students—students using the program were found to be on task 89 % of the time. Of this time, 7 % was spent in on-task conversations, confirming that students were interacting mathematically with peers and the teacher. This time on task is higher than time on task reported in traditional classrooms in suburban schools (Lee, Kelly, & Nyre, 1999), a finding which is particularly significant given that suburban students usually have higher time on task than students in urban schools.

⁹ This reflects a difference in emphases: Russian curricula and assessments—like those of other high-performing countries – emphasize depth of understanding of arithmetic and elements of algebra, while curricula and assessments in the United States are usually broad in their coverage of topics at the expense of depth (Schmidt et al., 2002).

Observations furthermore found that students experienced occasional delight (3 % of the time), low boredom, frustration, and confusion (10 %, 7 %, and 9 %, respectively), and a high rate of engaged concentration (71 %). It has not yet been investigated whether student affect in Genie 2 is associated with learning outcomes, but in other learning systems this has been the case (Baker et al., 2010).

Survey Results To determine student, teacher, and principal views of the system, Reasoning Mind staff administered anonymous surveys to samplings of 1148 students, 240 teachers, and 66 principals at the end of the 2011–2012 school year. The surveys showed that students in all grades preferred Genie 2 to traditional instruction. Furthermore, majorities of students in each grade said they at least “sort of” liked it. The overriding majority of teachers and principals accepted the program: 89 % of teachers and 93 % of principals wanted to continue using the program the next year. Teachers rated their overall experience with the program as +1.07 on a scale from –2 (very poor) to +2 (excellent). Of all individual components of the program, teachers gave the highest rating (+1.28) to the in-person support of implementation coordinators. Further confirming the importance of implementation coordinator support and in-person visits, 50 % of teachers felt they had the right number of visits from implementation coordinators, 31 % wanted more, and 19 % wanted fewer visits.

For details on the manner of survey administration, graphs of response data, and some discussion of the results, please refer to [Appendix 3: Student, Teacher, and Principal Survey Results](#).

General Principles for Cross-Cultural Curriculum Transfer Using Intelligent Tutoring Systems

The challenges encountered and lessons learned in the course of this work suggest several general principles for designing intelligent tutoring systems intended for cross-cultural transfer of curriculum and instructional methods. In this section, we list and discuss these principles. Taken together, the principles can be seen as a proposed methodology for cross-cultural curriculum transfer using intelligent tutoring systems. We will say “source country” to mean the country whose education system is being modeled, and “target country” to mean the country where the resulting intelligent tutoring system is being introduced.

1. **Design Knowledge Representation Structures Through Knowledge Engineering Educators in the Source Country.** It was learned in the course of this work that knowledge representation structures are specific to curricula—or, more concretely, to curriculum traditions. For example, a knowledge representation structure that uses the same object—in Genie 2’s terminology, the “objective” object—to refer to a knowledge component as to a unit of instructional content is less well-suited¹⁰ to modeling Russian teachers than a structure where these notions have separate

¹⁰ The architecture of Genie 2 combined these two notions into one, and this decision was made before enough knowledge engineering had been done to realize that it was suboptimal.

entities. This is because Russian mathematics course plans include a variety of lesson types, including lessons introducing new material, lessons meant to discuss connections between two or more previously covered topics, lessons meant to review some combination of previous topics in context of one another, and many other lesson types besides. This results in a complex, many-to-many relationship between topics (or knowledge components) and lessons. By contrast, American teachers usually organize instruction by knowledge components—with each unit introducing and then practicing a specific topic—and therefore their practices could be modeled more faithfully with a structure combining the two notions.

This point has several implications. First, it illustrates how the pedagogical model can be interconnected with the student model; Wenger (1987) argues that both models are necessary, and the present example shows that they must be designed in each other's context. Second, it suggests that modeling teachers from different traditions may be a method for uncovering additional classes of adaptive mechanisms. And third, it determines constraints on the generalizability of the Genie 2 platform to other nations' curricula. More specifically, modeling other nations' curricula through this kind of approach is not simply a matter of modifying the instructional content and the pedagogical rules: rather, the adaptive mechanisms and knowledge representation structures need to be customized for each curriculum tradition. This having been said, the degree to which they must be changed varies depending on how different the curriculum tradition is; for example, implementing a Singaporean curriculum—where, like in Russia, teachers take a primarily didactic approach—would be substantially easier than attempting to implement a Japanese curriculum. This is because in the Japanese approach, students work in small groups to find problem solutions, which are subsequently discussed with the teacher and the whole class (Stigler & Hiebert 2009). Capturing this group dynamic would require not only different knowledge representation structures, but also a fundamentally different student interface.

2. **Involve Teachers from the Source Country in the Development of Each Individual Piece of Content.** In the course of modeling the Russian teachers, it became apparent that they regarded the particulars of each topic as primary, and the application of general instructional principles as only secondary. (This is discussed in more detail in “Russian Mathematics Instruction.”) This implies that in order to successfully achieve cross-cultural transfer of these teachers' instructional methods, it is insufficient to enumerate instructional principles and then write content conforming to those principles: instead, the teachers themselves must be involved in the creation of all instructional content.

Some countries do not have as consistent a body of shared content-specific practices and mathematical knowledge for teaching as does Russia. For example, Stigler and Hiebert (2009) and Ma (1999) suggest that United States teachers do not have nearly as much such shared knowledge as do teachers in (respectively) Japan and China. While this may be interpreted as suggesting that this principle (namely, of involving source country teachers in all content development) is not always important, in practice it would only make sense to choose source countries with strong mathematics outcomes, and these source countries—as is certainly the case in Russia, Japan, and China—*do* tend to have a body of shared content-specific knowledge.

The application of this principle comes with a substantial content development burden. Because each piece of content must be written, illustrated, and programmed individually, the development of a single course takes a team of several dozen people 2 years. This imposes considerable expense for expanding to new grade levels, and also makes it harder to modify existing materials. It is likely that improvements in content development interfaces will allow for material to be developed more quickly in the future, but the process is still bound to remain very labor-intensive on account of the amount of knowledge engineering that must be done for each content component.

3. **Edit Content for Surface-Level Appropriateness to the Target country's Culture.** Given the premise of an international curriculum implementation, one might expect that differences in culture would affect the wording of content, thus making it hard to adapt materials. Indeed, it is certainly true that some story problems as written by the Russian teachers were not appropriate for American audiences: problems included splitting a telephone bill in a communal apartment, musk melons, and gathering mushrooms in the woods. Some problems were culturally inappropriate in more subtle ways, such as one which provided information on two twins' smoking habits and asked the student to find which of them would die sooner. There were some similar cases with content illustrations made by Russian artists; for example, a picture showing the Vitruvian Man made American teachers uncomfortable, and was consequently removed.

Nevertheless, addressing these issues of surface-level cultural differences is actually a matter of routine. In the case of Genie 2, because content had to go through translation and substantial editing regardless, it was enough to build cultural adaptation into that process. Story problems could all be easily changed to different scenarios without affecting the mathematical structure, and illustrations were reviewed using a set of guidelines listing the subjects and images to be avoided.

4. **Focus not Only on Content, but Also on Instructional Experiences** The Genie 2 system has a number of shortcomings in capturing certain aspects of Russian mathematics instruction; some of these shortcomings are described by Valcarce (2012). To list a few, Genie 2 fails to adequately model explicit coaching on mathematical solution writing, students' presenting of work to their peers, the integration of daily homework with planned lessons, and the use of periodic quizzes and exams to motivate retention of newly acquired knowledge.

These shortcomings were first noticed after the system was fully developed, put into use, and then analyzed in cooperation with the team of Russian teachers who participated in making it; they were caused by a failure to center knowledge engineering from the beginning around *instructional experiences* rather than purely content. By asking source country teachers to write content directly for an intelligent tutoring system, one immediately imposes constraints on what they can do. This can lead teachers to discard those components of their practice that do not obviously fit into the new genre. Furthermore, experts are not always equipped to model their own expertise, and this introduces an additional source of modeling error.

Thus, developers of intelligent tutoring systems for cross-cultural transfer of instructional methods would do well to begin by classifying the key instructional experiences of students in the source country, and only *then* designing learning interfaces and expected (offline) implementation practices meant to deliver these experiences to students in the target country.

In addition to this kind of front-end focus on instructional experiences, it is advisable to constantly evaluate the experiences of students using the system and to compare them against the experiences of source country students. One method found to be particularly helpful is to create an explicit catalog of source country instructional experiences (developed through interviews with expert teachers in the source country). Another helpful method is to bring source country expert teachers to the target country so that they can observe implementations of the system and provide their feedback on the specific instructional experiences students are (or are not) having.

5. Understand and Design for Existing Accountability Structures in the Target Country. The accountability structures of schools in different countries—and even in the same country at different times—can vary considerably, and this has very significant implications for whether and how schools will choose to implement intelligent tutoring systems. For example, United States schools are subject to the No Child Left Behind Act of 2001, a national law that imposes severe penalties on schools whose students fail to make progress on state-administered standardized achievement tests (Hamilton et al., 2007). By contrast, Russian schools have much more limited external accountability, but at various times in their history have been subject to national laws mandating the use of centrally approved curricula and lesson sequencings (Karp & Vogeli, 2010).

In virtually all of the school districts using Genie 2, administrators had responded to the requirements of No Child Left Behind by implementing extensive yearlong testing programs to monitor students' learning. These programs usually took the form of frequent "benchmark" tests, substantial exams consisting of problems meant to emulate those of the high-stakes test. Each benchmark test covered the topics scheduled to be taught in the preceding one or two months. By the end of the first semester, students were often expected to have covered all of the tested topics, freeing the second semester for review and additional practice.

These benchmarks imposed a *de facto* expected sequencing of topics throughout the year. These sequencings were often mathematically arbitrary, and were different in every district. Most of the teachers and administrators involved in using the program assumed that mathematical topics *could* be taught in an arbitrary order, and had the concomitant expectation that objectives in Genie 2 could and would be reordered to match the district's chosen sequencing. However, this was impossible to do, as topics in Russian mathematics courses are sequenced to form a coherent progression (cf. Schmidt, Houang, & Cogan, 2002), and reordering them would disrupt this coherence.

This mismatch between districts' expectations and the system's requirements posed a substantial obstacle to the program's expansion. First of all, many districts declined to even begin implementations of the program when they learned that topics must necessarily be covered in a different order from the district's sequencing. In districts that did choose to implement the program

despite the fixed sequencing, the benchmarks often posed a threat to implementation fidelity: when students using the program did worse on a benchmark than other students (which was bound to happen, given that the benchmarks were aligned to the sequence studied by the students not using the program), there was a strong pressure to reduce time online and switch to test preparation to obtain better results on the next benchmark.

One school district developed two sets of benchmarks, one aligned to the Reasoning Mind sequencing and the other to the regular district sequencing, and administered both sets to all students. As expected, each group of students consistently did better than the other on the benchmarks aligned to its own sequencing.

Thus, the issue of sequencing complicated the program's expansion into new districts and compromised the quality of implementations. Some methods have been developed to mitigate—though not altogether remove—the issue: focusing on district-wide adoptions (so that students using the program are not subjected to monthly comparisons using misaligned instruments), consistently explaining the motivation for the fixed sequencing to administrators, and supporting teachers in doing some amount of test preparation concurrently with students' use of the program. With the most committed partner districts, it has been possible to develop benchmark tests aligned with the program's sequencing and completely replace the previous benchmarks with these new tests; while difficult to accomplish, this is much closer to a complete solution to the problem.¹¹

Therefore, the accountability systems of both the source country and especially the target country have serious implications for adoption and implementation. These accountability systems should be carefully studied in designing cross-cultural intelligent tutoring systems and should inform decisions concerning system design and implementation.

6. Provide Considerable In-Person Technological as Well as Instructional Support to Teachers in the Target Country. Because so many aspects of the program were new to teachers—the different classroom setup, the curriculum, the teacher interface, and so on—considerable effort was required to provide sufficient training, monitoring, and support. In particular, frequent in-person school visits have been essential for giving teachers feedback and encouragement. While this has substantially increased the cost to schools (creating an obstacle to expansion), it has led to strong relationships with participating teachers and has turned out to be essential for strong implementation of Genie 2.

This kind of instructional support is best developed iteratively: i.e., through carefully studying existing implementations, cataloging deficiencies, and then redesigning the teacher training and support model to address them for future years.

¹¹ It should be noted that alignment of the curriculum—i.e., not the sequence, but the total scope of what is covered during the year—is an easier task. Russian curricula heavily overlap with United States standards (most notably, the new Common Core State Standards), so that the topics included in the standards but not covered by the corresponding Russian grade levels are not numerous, and can be addressed through supplemental materials, whether online or provided in the form of offline lessons to be delivered by teachers. The problem remains of familiarizing students with the format of the specific end-of-year high-stakes test, but this can be partly addressed by giving students practice problems; there is a special learning mode in Genie 2 for this purpose (called the Test Mode).

In addition to enabling stronger implementations, in-person support greatly increases the feedback obtained from target country teachers using the program. Such feedback is essential for iterative refinement of the system, as any educational technology will only be implemented to the extent that it meets the needs and constraints of teachers (Cuban, 1986), and technologies based on other countries' education systems are at particular risk of unintended and unnoticed misalignment with these needs and constraints.

Of course, providing this kind of in-person support is expensive, and therefore is only practical on a large scale in countries where school budgets are sufficiently large to cover the costs. The United States is such a country: while the development of software and curriculum has been subsidized by philanthropy, schools using Reasoning Mind have covered the cost of providing in-person support.

- 7 Evaluate Implementation Fidelity in the Target Country Using Unambiguous and Reliable Protocols. Because cultural expectations for instruction in the source and target country can be very different, the teachers implementing a system may not realize when they are not using it as intended. Furthermore, even a trained school support staff—if consisting of nationals of the target country—may not understand what components are missing or misconstrued in a given implementation.

In Genie 2, for example, it was found that United States teachers did not ascribe as much value to careful note-taking and showing of written work as did Russian teachers, did not always understand the importance of remediating gaps in prior knowledge, and deprioritized solving more advanced (level B and level C) problems. These differences in mathematical values were addressed by implementation coordinators, who observed teachers' classrooms and gave them feedback. To ensure that the observations were consistent and achieved the intended purpose, it was essential to develop clear protocols for observing and giving feedback (namely, the Implementation Levels described in "Implementing Genie 2 in Schools").

Discussion

Concerning the Effectiveness of Genie 2 From the point of view of student learning outcomes, the evidence indicates that students using the program learn more than students in traditional classrooms. The same is true of other intelligent tutoring systems; for example, Cognitive Tutors in algebra have shown at-scale effects of between 0 and 0.3 standard deviations on standardized tests (Ritter et al., 2007; Campuzano et al., 2009; Pane et al., 2013) and 0.7–1.0 standard deviations on more proximal instruments (VanLehn, 2011; Anderson et al., 1995; Koedinger et al., 1997), and there are encouraging results for systems such as Wayang's Outpost (Arroyo et al., 2004) and Assistments (Razzaq et al., 2005). It is important to note that the approach taken in Genie 2 is not an *alternative* to other approaches, as future learning systems could synthesize different lines of attack.

It is unknown how much the central idea of Genie 2—namely, modeling the Russian mathematics teaching approach—contributes to the system's success in engaging students and improving outcomes. However, there are several ways in which it is particularly likely to play a role. First, by organizing knowledge around mathematically

meaningful structures, teaching and employing explicit rules and definitions, and providing explanatory feedback, the curriculum could encourage students to develop deeper knowledge of mathematics. Second, through consistent and deliberately planned spaced practice of skills, the system could lead to greater procedural fluency. Third, through the use of guided instruction, fading, and the absence of mathematically meaningless tasks, the system could lead to more efficient use of instructional time, and this in turn could lead to greater learning. Fourth, by breaking concepts into constituent components and using mathematically precise language, the system's explanations could be clearer for students (indeed, students often report that they can understand the system's explanations better than those they receive in traditional math classes), and this could contribute to the observed low boredom and low time off task; this in turn has significant implications for outcomes: for example, Anderson & Schunn (2000) argue that time on task is essential to successful learning, and could account for much of the difference in outcomes between students in the United States and mathematically high-performing countries. And fifth, by immersing United States teachers in a stronger curriculum and providing them with support and professional development focused on mathematics content knowledge, the approach could increase the quality of instruction they provide to students in their classrooms; teachers' mathematical knowledge for teaching is associated with student learning (Hill, Rowan, & Ball, 2005). Further investigation is needed to determine which—if any—of these possible causes are responsible for the observed learning gains.

Computer and Human Components of the “System” Regardless of which causes are responsible, it has been clear since the 2005–2006 school year that robust teacher support is an essential precondition for successful implementation of Genie 2. Put another way, the “system” really includes not only the software platform and content, but also all of the procedures in place for training and supporting teachers and principals.

Viewed from this perspective, Genie 2 consists of a large number of components—an adaptive rule-based algorithm, instructional content, suitably trained and supported classroom teachers, online tutors—collectively designed to emulate the behavior of expert Russian mathematics teachers, or—more specifically—to recreate the instructional experiences that students in those Russian teachers' classrooms would have. Closed-ended tasks, such as providing explanations, checking numerical answers, and tracking data are performed by the computer system; meanwhile, teachers (and in some cases online tutors) perform tasks which are more open-ended or require physical presence in the classroom, such as managing the class, evaluating the quality of written work, providing interventions, encouraging productive peer interactions, and creating a positive class culture. This is much easier for teachers than implementing a foreign curriculum in a traditional classroom, since the computer system not only takes care of certain tasks requiring particularly deep curriculum-specific content knowledge (such as providing explanations of concepts and selecting appropriate problems), but also scaffolds all instructional experiences in the classroom, effectively immersing not only students, but also the teacher in the new curriculum.

Future Directions Genie 2 successfully captures a number of aspects of expert Russian teachers' practice. However, there are also several significant practices that it fails to

model. First, because the expert Russian teachers were instructed to write scripts directly for the computer implementation, it is to be expected that they discarded—either intentionally or without realizing it—those practices which they could not easily fit within the prescribed boundaries of the system. Second, the self-paced nature of the system and the organization of content in the system around “objectives” (each of which discusses a single concept or set of related concepts) is at odds with Russian course plans, which consist of 45-min lessons, each designed to be taken in one sitting and containing a wide range of different activities, including explanations of new material, review of old concepts, and “consolidation” of recently acquired ones; this means that every student studying in the system effectively receives a different “lesson plan” in each session, and these lesson plans do not correspond in any precise way to Russian curriculum lesson plans. Because the structure and storyline of each lesson is an important aspect of Russian pedagogical thinking and practice (Karp & Zvavich, 2011), this difference is significant. Third, in the Vygotskian tradition, Russian teachers give great importance to dialog in lessons: the teacher frequently calls individual students to the blackboard to participate in structured discussions with the teacher, and the rest of the class observes and learns from this dialog; this instruction through vicarious learning is not in any way captured in Genie 2. For a further discussion of the ways in which Genie 2 succeeds and fails in simulating Russian teachers, see M. Valcarce’s manuscript on the subject (Valcarce, 2012). The pedagogical model of Genie 2 could likely be strengthened by incorporating these and other currently missing aspects of the expert teachers’ practice.

Conclusion

The results presented here show that it is possible to use intelligent tutoring systems to implement one country’s mathematics curriculum and instructional practices in another country in a way that has high acceptance among students and teachers, and furthermore can be sustained at a scale of at least tens of thousands of students. As a practical contribution, we described Genie 2, a concrete such system which has successfully brought many aspects of Russian mathematics education to United States students; previously published works have found that this system improves student learning outcomes. Finally, we have outlined a collection of design principles extracted from the experience of designing, implementing, and iteratively refining Genie 2. Taken together, these principles represent a proposed methodology for the design of intelligent tutoring systems intended for cross-cultural transfer of curriculum and instructional methods.

Acknowledgments We would like to thank the many people—over a hundred software engineers, curriculum developers, teachers, artists, designers, school support staff, and many others—who have contributed to building and supporting the systems described here. We would also like to acknowledge Reasoning Mind’s partner schools, who have put great effort into successfully implementing the program. The development of the program was made possible by over 200 philanthropies and individual donors. We also thank Richard Mayer for first encouraging us to publish our findings from Genie 2, Ryan Baker, Alexander Karp, Maria Klawe, and Bill Schmidt for helpful conversations, Vincent Alevan for his guidance during revision, the anonymous reviewers for their thoughtful and detailed suggestions—which greatly improved the article—and Ryan Baker for his exceptionally valuable comments on the manuscript.

Appendix 1: History and International Reception of Russian Mathematics Education

The evidence mentioned in the beginning of the “Russian Mathematics Instruction” section suggests that elite mathematics outcomes in Russia—namely, the production of top research mathematicians and school-level contest winners—are first-rate, while general mathematics education outcomes are strong but somewhat weaker than in the Pacific Rim countries, which outperform Russia on the TIMSS assessments.

While this evidence is already sufficient to justify an interest in Russian mathematics education methods, it likely understates the strength of the Russian mathematics curriculum for three reasons. First, because of the strong cultural emphasis that Pacific Rim countries place on school education, many students in these countries attend after-school education programs (Stevenson, 1994). While similar patterns can be observed in Russia’s largest cities (Moscow and St. Petersburg), this is not broadly true of Russian education. Therefore, the Pacific Rim countries have educational strengths outside of curriculum that contribute to their performance.

Second, strong general education is not the only contributor to strong performance on international tests; in particular, test-taking skills play a substantial role in test performance, placing countries with entrenched high-stakes testing systems (such as the United States, Singapore, and Korea) at a relative advantage. Russian students do not take high-stakes tests until high school, and until the last several years, there have been no high-stakes tests—and in fact almost no multiple-choice tests of any kind—in Russian schools.

And third, while the Pacific Rim countries have seen a strengthening of their mathematics education systems in the last decades, Russia has seen a decline. There has been a great exodus of mathematics teachers and researchers—for example, of the seven Russian Fields medalists mentioned previously, six now reside outside of Russia. Furthermore, whereas in Soviet times good teachers were to be found throughout the country (Stefimova, 2010), the quality of entrants to pre-service teacher education programs has since then declined sharply (Yu. Konovalova, personal communication, July 2, 2013). Thus, while the curriculum has largely remained the same—and there are still teachers who teach it as effectively as in the 1970s and 1980s—the overall education system has worsened, and with it student outcomes.

So, while an analysis of overall outcome measures in mathematics education is a good departing point for assessing the strength of a curriculum, it does not account for all of the factors. For this reason, it will be helpful to complement this analysis with a consideration of the background of the Russian curriculum, along with the main influences and events in its history.

The modern age of Russian mathematics education dates from the reign of Peter the Great. As Russia’s first imperial tsar, Peter invested heavily in mathematics education to support his military endeavors. This began a centuries-long history of government support of mathematics and mathematics education. A particularly important role was played by the great mathematician L. Euler, who not only wrote a number of textbooks in basic mathematics—

including texts in arithmetic and algebra—but also created a school of mathematics curriculum development which remained influential for over a century. It was Euler who first articulated the necessity of both mathematical precision and accessibility, an approach which has remained a defining characteristic of Russian mathematics curriculum to this day (Polyakova, 2010).

Euler's example was followed by Russian and subsequently Soviet mathematicians, who maintained a constant involvement with school mathematics. Lobachevsky wrote school textbooks, and Chebyshev developed national pedagogical guidelines (in today's language, "standards") to update the curriculum in the mid-19th century. Around the turn of the 20th century, the teacher and curriculum developer A.P. Kiselev published three textbooks—on arithmetic, algebra, and geometry—that both incorporated elements of the preceding tradition and became the foundation for Russian curricula used from the 1930s onward. Kiselev's geometry textbook—which has been translated into English by A. Givental—remains in use in some schools to the present day. In the 1960s, another important development took place: the mathematician N. Kolmogorov led a series of mostly successful curriculum reforms. These reforms sought to modernize the curriculum, and produced a large number of influential textbooks (Abramov, 2010).

The method of development of the textbooks produced during the Kolmogorov reforms has much to do with their success: mathematicians, psychologists, curriculum researchers, and practicing teachers were all involved in the process. Teams wrote textbooks together with detailed lesson plans; subsequently, the teachers on each team used the textbooks and lesson plans in their classes, bringing outcome data back to the whole team to discuss and iteratively improve the materials. Some of the most successful books went through as many as four separate year-long iterations before approval and distribution to the nation's schools. These textbooks included an elementary school program by M. I. Moro et al.¹² as well as a program for grades 5 and 6 (originally 4 and 5) by N. Ya. Vilenkin et al.; these books remain the most widely used textbooks in the corresponding grade levels to the present day (Abramov, 2010).¹³

While in Soviet times, schools were frequently required to use the same textbook across the country, modern Russian schools are allowed to select from a short list of approved textbooks. A number of strong textbooks have been developed since the fall of the Soviet Union, including textbooks by L. G. Peterson and B. P. Geydman.

Russian curriculum development in Soviet times was heavily influenced by Soviet psychology, and especially by the Vygotskian tradition. Aspects of this—such as a focus on "the general development of the child" and an effort to

¹² This curriculum has been translated into English by the University of Chicago School Mathematics Project and is available in the United States.

¹³ It should be noted that not all of the textbooks produced within the scope of the Kolmogorov reforms were successful. For example, the geometry textbook authored by Kolmogorov himself, while exceptionally innovative in its approach, proved too theoretically demanding and fell out of use a decade after its introduction (Abramov, 2010).

calibrate the difficulty of challenges to the level of learners—can be found in the textbooks of the Kolmogorov reforms and in the attitudes and priorities of Russian teachers. One curriculum—the Elkonen-Davydov curriculum—was developed entirely within the Vygotskian tradition; while it is not considered mainstream, it is well-regarded by Russian teachers. Interestingly, this curriculum has attracted interest from researchers in the United States, and has even been used in an American school (Schmittau & Morris, 2004).

The Russian curriculum has influenced the curricula of several other countries, and especially the Eastern Bloc countries. In his textbook for mathematics teachers, Milgram (2005) suggests that the Russian curriculum was adopted by China in the 1950s, and from there became the foundation of the Singaporean curriculum. While the authors have not been able to independently confirm this provenance of the Singaporean curriculum, it does share a great deal in common with Russian programs, as Milgram argues. Milgram writes:

The best foreign programs in mathematics are so overwhelmingly successful with such a high percentage of students that we can learn a great deal... by studying how these programs handle the core topics and their development. Foremost among these programs is the Russian mathematics program that was developed during the 1930's and 1940's. It was used in Israel from the late 1940's to the early 1970's during which time their outcomes were the best in the world. It was adopted by China during the 1950's, and from China was modified for use in Singapore. It is also the core program underlying the current programs in former iron countries such as Bulgaria, Hungary, Poland, and Romania. Thus, we pay a great deal of attention to the way in which the Russian program develops the core concepts in mathematics during the early years, and we also reference the Singapore program extensively to learn about how the... topics above are treated in countries where instruction in mathematics is successful.

Throughout his textbook, Milgram refers to the M. I. Moro textbook to illustrate a number of important aspects that this book shares with other Russian texts, including an attention to mathematical precision, the early introduction of algebraic notation and concepts, and an emphasis on problem-solving. As such, Milgram's book is a good reference for English-speaking readers wishing to familiarize themselves with the main attributes of Russian curriculum in the lower grades.

Milgram is not alone in his high regard for Russian curriculum materials, as many curriculum specialists outside Russia consider them to be exemplary (W. Schmidt, personal communication, January 10, 2013). For a view from the other side, Russian mathematician A. Toom has provided in his essay “A Russian Teacher in America” (Toom, 1993) an amusing account of his experiences in the United States; in the essay, curriculum is discussed among a variety of other aspects of the two education systems.

In conclusion, while the Russian curriculum is not the only successful mathematics curriculum, it is by many accounts at least *among* the strongest, and thus an appropriate choice for cross-cultural adoption.

Appendix 2: Two Excerpts from Lesson Scripts

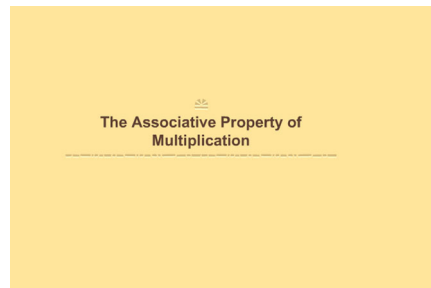
Here, we present two excerpts from Genie 2 lesson scripts. These excerpts serve to demonstrate what typical lesson scripts and the resulting content are like. Additionally, they give some idea of the variety of methods available to content authors. The excerpts are given with the script on the left and a sequence of selected corresponding screenshots from the final material on the right. They are given in the final form: i.e., not in the original (more skeletal) version written by the Russian teachers, but rather in the detailed and complete version arrived at following editing by the curriculum development team.

The first excerpt presents the associative property of multiplication, and is a small piece (about 10 %) of an objective entitled “Properties of Multiplication.” The other sections of this objective discuss the commutative property, multiplication by powers of 10, multiplication by 0 and 1, and multiplication by round numbers (i.e., numbers ending in one or more zeros). The material is intended for 4th or 5th grade students.

[Chapter 3]

The Associative Property of Multiplication

[SUBTITLE]



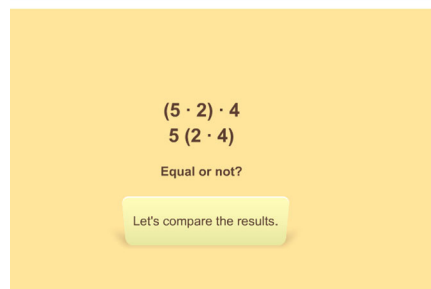
Animation 4

[Preview:]

$(5 \cdot 2) \cdot 4$
 $5(2 \cdot 4)$
 Equal or not?

[Next->]

Let's compare the results.



[Next->new screen]

Let's find $(5 \cdot 2) \cdot 4$.

[Next-> remove "Let's find" keep the expression; add an arc below $5 \cdot 2$, mark it with 10, then copy the expression with the arc, move 10 up to replace $(5 \cdot 2)$ in RHS; finish the expression with $=$ [uINT]¹⁴:]

$$\underbrace{(5 \cdot 2)}_{10} \cdot 4 = 10 \cdot 4 = \text{[uINT]}$$

[Check answer. Correct is 40. If incorrect, cross out incorrect answer; Next-> replace the incorrect answer with the correct one.]

[Next->keep the expression on top; add:]

Let's find $5(2 \cdot 4)$.

[Next->remove "Let's find," keep the expression; add below $2 \cdot 4$ an arc marked 8; continue the expression with $=$ 5·[uINT]"]

[Check answer. Correct is 8. If incorrect, cross out incorrect answer; Next->replace the answer window with the copy of the 8 from below the arc.]

[Next-> finish the expression with $=$ [uINT]"]

[Check answer. Correct is 40. If incorrect, cross out incorrect answer; Next->replace the incorrect answer with the correct one.]

[Next->INFLATE both 40, add:]

So, $(5 \cdot 2) \cdot 4 = 5(2 \cdot 4)$.

[Next->in "remember" box]

Changing the grouping of the factors does not change the product.

¹⁴This means "integer entry," and is the most common of the various formats in which students enter answers in Theory animations.

[Next->new screen, keep the rule "changing..."; add:

Actually, this is true for any three numbers. We can use letters to write this property.

[Next->add on screen:]

$(a \cdot b) \cdot c$ [add an equals sign; copy to the right, then move the parentheses in RHS to make]

$$= a(b \cdot c)$$

[Next->]

This is called the associative property of multiplication.

[Next->Genie:] Remember, <student's name>, we often omit the multiplication sign between letters.

[END OF ANIMATION]

Remember

Changing the grouping of the factors does not change the product.

Actually, this is true for any three numbers. We can use letters to write this property.

$$(a \cdot b) \cdot c = a(b \cdot c)$$

This is called the associative property of multiplication.

Static Screen 5

[Standard WRITE THIS DOWN screen with text:]

Associative Property of Multiplication
 $(ab)c = a(bc)$

[END OF STATIC]

WRITE THIS DOWN

Associative Property of Multiplication

$$(ab)c = a(bc)$$

Animation 6

[Preview:]

$$(ab)c = a(bc)$$

[Next->]

The location of the parentheses does not affect the result. So, we can remove the parentheses.

[Next->add: $(ab)c = a(bc)$ below; remove with effect parentheses in both expressions and combine them into one expression abc .]

[Next->move "abc" to make on top:]

$$(ab)c = a(bc) = abc$$

[END OF ANIMATION]

The location of the parentheses does not affect the result. So, we can remove the parentheses.

$$(ab)c = a(bc) = abc$$

Exercise 7

[On the screen are 3 unpainted balloons with expressions " $(3 \cdot 17) \cdot 17$ "; " $11(7 \cdot 17)$ "; " $(13 \cdot 17) \cdot 7$ and a tin of paint. The student can use the paint on one of the objects.]

Choose an expression equal to $13(17 \cdot 7)$, and paint that balloon.

[Check answer. Correct is $(13 \cdot 17) \cdot 7$. If incorrect, cross out incorrect answer; Next->show the correct answer]

[END OF EXERCISE]

Choose an expression equal to $13(17 \cdot 7)$, and paint that balloon.

Animation 8*[Preview:]*Let's find the product $5(2 \cdot 8)$.*[Next->]*

Let's use the associative property of multiplication so we can multiply 5 and 2 first to get a round number, and then multiply the result by 8.

[Next->new screen]

[On top: " $5(2 \cdot 8)$ " and "Let's use the associative property of multiplication."]

[Next->add below:]

$$5(2 \cdot 8) = (5 \cdot 2) \cdot 8$$

[Next->continue the expression with]

$$= [\text{uINT}] \cdot 8$$

[Check answer. Correct is 10. If incorrect, cross out incorrect answer; Next->add an arc below $(5 \cdot 2)$, mark it with 10; copy the 10 to replace the answer window.]

[Next-> finish the expression with] = [uINT]

[Check answer. Correct is 80. If incorrect, cross out incorrect answer; Next-> replace the incorrect answer with the correct one.]

*[END OF ANIMATION]*Let's find the product $5(2 \cdot 8)$.

Let's use the associative property of multiplication so we can multiply 5 and 2 first to get a round number, and then multiply the result by 8.

 $5(2 \cdot 8)$

Let's use the associative property of multiplication.

$$5(2 \cdot 8) = (5 \cdot 2) \cdot 8 = \square \cdot 8$$

 $5(2 \cdot 8)$

Let's use the associative property of multiplication.

$$5(2 \cdot 8) = (5 \cdot 2) \cdot 8 = 10 \cdot 8 = \square$$

Exercise 9

Calculate. (Use the associative property of multiplication.)

$$2(5 \cdot 7) = [\text{uINT}]$$

[Check answer. Correct is 70. If incorrect, cross out incorrect answer;]

[RESPONSE 1]

[Next->remove the answer window; continue the expression with copy $2(5 \cdot 7)$; then replace the parentheses in the RHS to make $(2 \cdot 5) \cdot 7$]

[Next->continue the expression with:]
 $= 10 \cdot 7$ [pause] = 70

[END OF RESPONSE 1]

[Next->new screen]

Calculate. (Use the associative property of multiplication.)

$$(12 \cdot 2) \cdot 5 = [\text{uINT}]$$

[Check answer. Correct is 120. If incorrect, cross out incorrect answer.]

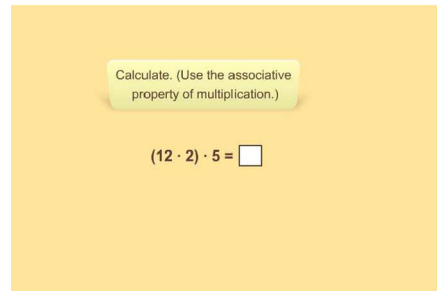
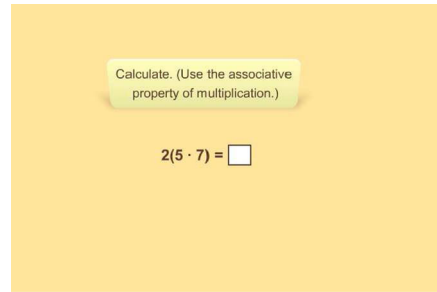
[RESPONSE 2]

[Next->remove the answer window; continue the expression with copy $(12 \cdot 2) \cdot 5$; then replace the parentheses in the RHS to make $12(2 \cdot 5)$]

[Next->continue the expression with:] =
 $12 \cdot 10$ [pause] = 120

[END OF RESPONSE 2]

[END OF EXERCISE]



This concludes the first excerpt. The second excerpt covers basic geometry, and is intended for 2nd or 3rd grade students. It is an exercise given to students who have just learned about naming line segments and about “segment chains” (i.e., piecewise linear curves). Among other things, this exercise gives an example of branched logic in Theory material: nested questions with system responses that depend on the particular answers submitted by students. The Math Pirate character appearing in this animation was inspired by the late University of Chicago mathematician and mathematics educator Paul Sally, Jr., to whose memory this article is dedicated.

Exercise 3

[Desk with a sheet of paper – a drawing of a segment chain; its vertices are labeled, but they are too small to see at this point. The Math Pirate stands near the desk with a bubble: “I tricked the Naming Machine, and I bet I can trick <Student name> too!”]



[Next-> Pirate:]
I bet you’ll never guess the names of these line segments, <Student name>!



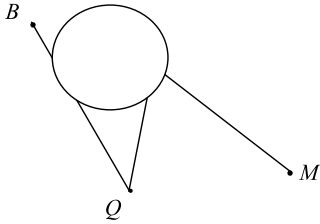
[Next-> He puts his compass on the drawing.]



[Next-> New screen. Show the drawing; this time, it’s larger and the letters are large enough to read, except the one covered with the compass. Part of \overline{BQ} is also hidden:]

The Math Pirate marked and labeled points B , C , M , and Q . He also drew line segments \overline{BQ} , \overline{QC} , and \overline{MC} .

Click on line segment \overline{QC} .



[Text:]

The Math Pirate marked and labeled points B , C , M , and Q . He also drew line segments \overline{BQ} , \overline{QC} , and \overline{MC} . Click on line segment \overline{QC} .

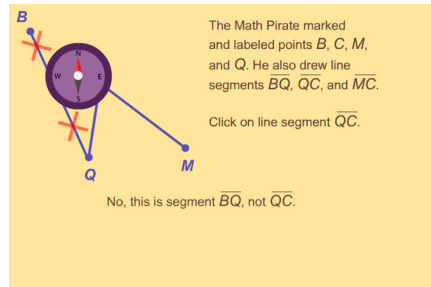
[Only allow clicking on visible parts of segments (not too close to their endpoints). Upon click, the visible part of the segment changes its color to a very bright one and becomes slightly thicker. Until SUBMITting, the user can change her mind and select another segment; the previous choice is unhighlighted then.]

[Check answer. Correct is \overline{QC} (see drawing at the end of the exercise).]

[If incorrect, cross out incorrect answer (the visible part of the segment); change the segment's color and thickness back to original.]
[RESPONSE]

[If the user clicked on \overline{BQ} ,

[Next->] No, this is segment \overline{BQ} , not \overline{QC} .

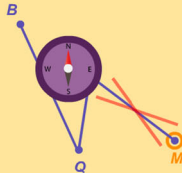


[If the user clicked on MC.]

Next-> CIRCLE point M together with its label.

[Next->] M is an endpoint of the segment you chose.

So, this segment has M in its name. [Next->] Therefore, it's NOT \overline{QC} .



The Math Pirate marked and labeled points $B, C, M,$ and $Q.$ He also drew line segments $\overline{BQ}, \overline{QC},$ and $\overline{MC}.$

Click on line segment $\overline{QC}.$

M is an endpoint of the segment you choose. So, this segment has M in its name. Therefore, it's NOT $\overline{QC}.$

[In case of any incorrect answer, continue with]

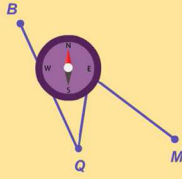
[RESPONSE 2]

[Next-> New screen: drawing from the previous screen is on top; text:]

[QUESTION]

The Math Pirate marked and labeled points $B, C, M,$ and $Q.$ He also drew line segments $\overline{BQ}, \overline{QC},$ and $\overline{MC}.$

[Next->] Try again: click on line segment $\overline{QC}.$



The Math Pirate marked and labeled points $B, C, M,$ and $Q.$ He also drew line segments $\overline{BQ}, \overline{QC},$ and $\overline{MC}.$

Try again: click on line segment $\overline{QC}.$

[Check answer. Correct is \overline{QC} (see drawing in the end of the exercise).]

[If incorrect, cross out incorrect answer (the visible part of the segment); change the segment's color and thickness back to original.]

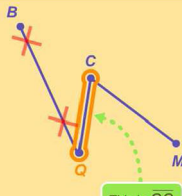
[Next->] One of the endpoints of \overline{QC} is $Q.$

[Next-> CIRCLE Q in the drawing. Next->

Highlight visible parts of \overline{QC} (color to be chosen be designer/ animator).

Next-> Point an arrow to \overline{QC} ; the arrow is labeled "This is $\overline{QC}.$ "

[Next-> Remove the compass; The entire



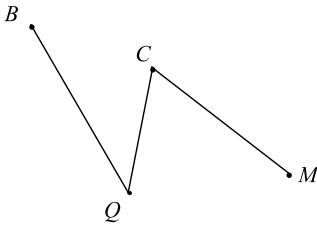
The Math Pirate marked and labeled points $B, C, M,$ and $Q.$ He also drew line segments $\overline{BQ}, \overline{QC},$ and $\overline{MC}.$

Try again: click on line segment $\overline{QC}.$

This is $\overline{QC}.$

One of the endpoints of \overline{QC} is $Q.$

drawing is:]

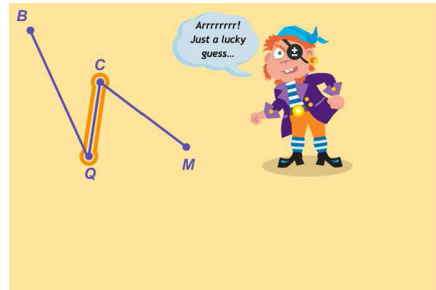


[If correct:]

[REWARD RESPONSE:]

[Next-> Remove the compass; The entire drawing is shown below. QC is shown highlighted (by the user). Pirate appears.]

[Next-> Pirate:]
Arrrrrrrr! Just a lucky guess...



[Next-> Pirate:]
I'll trick you next time,
<Student name>!
[END OF REWARD RESPONSE]

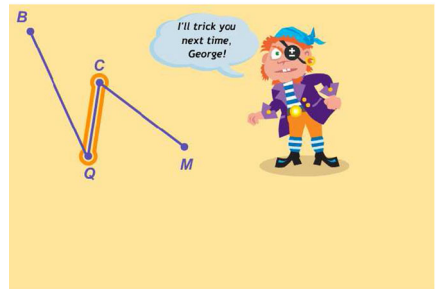
[END OF QUESTION]

[END OF RESPONSE]

[If correct:]

[Show REWARD RESPONSE (described above).]

[END OF EXERCISE]



Appendix 3: Student, Teacher, and Principal Survey Results

Anonymous surveys were administered to samplings of 240 teachers and 66 principals at the end of the 2011–2012 school year. Districts surveyed were representative of the program's participants, and included rural and urban districts of varying sizes (from small to large) in Texas, California, Louisiana, West Virginia, and Oklahoma. In each surveyed district, all participating principals were sent a survey. Some districts required principals to consent before teachers could be surveyed; in districts without this requirement, all teachers received surveys, while in districts with this requirement, surveys were sent to all teachers whose principals consented. Approximately half of teachers and principals who received surveys responded.

Of the individual components of the program, teachers gave the highest rating (+1.28) to the in-person support of implementation coordinators. Teachers also found the reports to be useful. Teachers rated the reports and teacher interface usability to be +1.04 (on a scale from -2 to +2), and 85 % of teachers said that they used reports at least once a week, with 66 % saying they used them more frequently than that. The four most popular reports were used by 57 %, 58 %, 70 %, and 82 % of teachers.

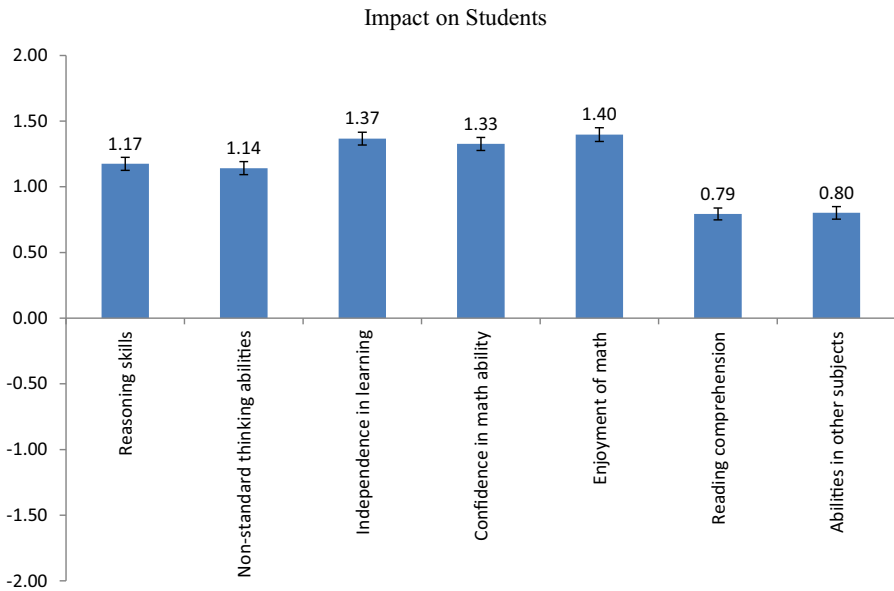


Fig. 11 Teacher responses to end-of-year survey

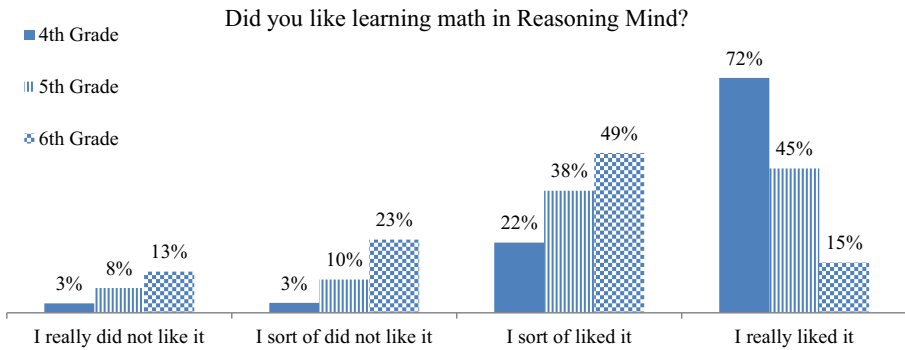


Fig. 12 Student attitudes towards the program

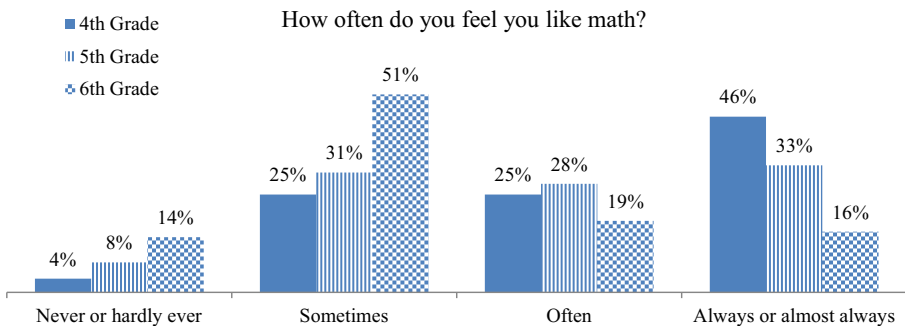


Fig. 13 Student attitudes towards mathematics

Would you prefer to learn math in Reasoning Mind or in a regular classroom?

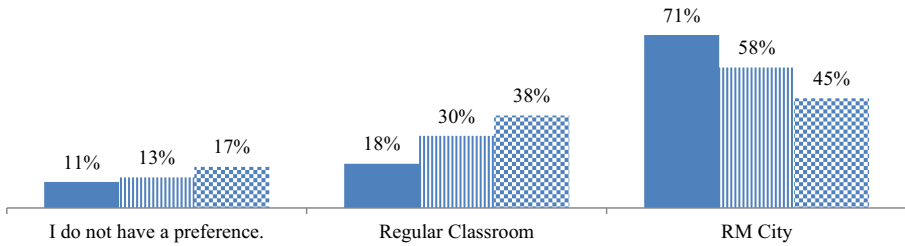


Fig. 14 Student preferences

Teachers were also asked whether they felt the program impacted students in a number of areas. Figure 11, shows their responses (on a scale with -2 being “significantly regressed” and $+2$ being “significantly improved”), with error bars for a 95 % confidence interval.

Students were surveyed in a similar manner; the districts that gave permission were two mid-to-large urban districts in California, two mid-to-large urban districts in Texas, and one Texas private school serving economically disadvantaged students. Students were surveyed in all schools where the principal gave permission. Some districts required additional parental permission, substantially lowering response rates in these districts; altogether, about half of eligible students in participating districts could be surveyed, for a total of 1148 respondents. Only students in grades 4–6 were surveyed, since younger children are substantially more likely to misinterpret questions and therefore can give less reliable responses (Borgers, de Leeuw, & Hox, 2000).

The results (Fig. 14.) show that students in all grades preferred Genie 2 to traditional instruction. Majorities of students in each grade said they at least “sort of” liked it (Fig. 12.). However, students liked the program less with each successive grade. This might be explained by a number of factors: students generally like mathematics less as they enter adolescence (Midgley, Feldlaufer, & Eccles, 1989), a finding which is apparently corroborated by Fig. 13; the 4th grade curriculum in Genie 2 is the most recently developed one, and consequently is more interactive and animated than the others; the 5th and 6th grade curricula are core curricula, and therefore may be more likely to be perceived as hard work, whereas the ancillary 4th grade might be perceived as a change of pace from traditional classwork; the interface is more age-appropriate to

Points and streaks are the main reason I do math in Reasoning Mind.

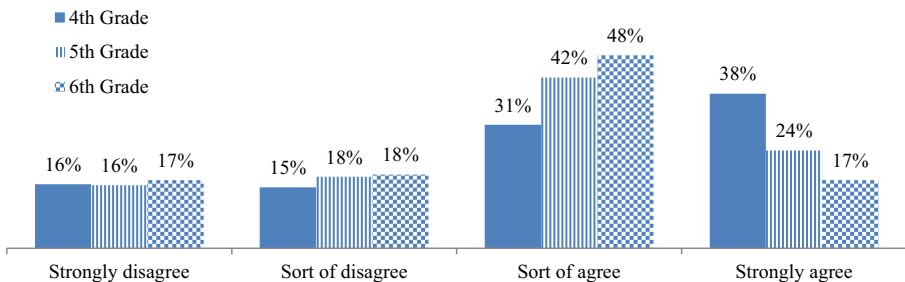


Fig. 15 Student valuation of extrinsic motivators

younger students; and finally, the 6th grade curriculum is particularly difficult for students owing to its dry presentation and demanding mathematical language.

In every grade surveyed, most students agreed that extrinsic rewards were their main motivator in the program (Fig. 15.). However, students in higher grades were more likely to “sort of agree” with this statement than to “strongly agree.”

References

- Abramov, A. (2010). *Toward a history of mathematics education reform in Soviet schools (1960s–1980s)* (pp. 87–140). History and World Significance: Russian Mathematics Education.
- Adams, D., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & Van Velsen, M. (2012). Erroneous examples versus problem solving: can we improve how middle school students learn decimals? In N. Miyakem, D. Peebles, & R. P. Coppers (Eds.), *Proceedings of the 34th meeting of the cognitive science society (CogSci 2012)* (pp. 1260–1265). Sapporo: Cognitive Science Society.
- Andersen, E., Liu, Y. E., Snider, R., Szeto, R., Cooper, S., & Popović, Z. (2011, June). On the harmfulness of secondary game objectives. In *Proceedings of the 6th International Conference on Foundations of Digital Games* (pp. 30–37). ACM.
- Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. *Advances in instructional psychology*, 5, 1–34.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4(2), 167–207.
- Arroyo, I., Beal, C., Murray, T., Walles, R., & Woolf, B. P. (2004, January). Web-based intelligent multimedia tutoring for high stakes achievement tests. In *Intelligent Tutoring Systems* (pp. 468–477). Springer Berlin Heidelberg
- Baker, R. S., D’Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners’ cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223–241.
- Baroody, A. J., & Ginsburg, H. P. (1983). The effects of instruction on children’s understanding of the “equals” sign. *Elementary School Journal*, 84, 199–212.
- Behr, M., & Post, T. (1992). Teaching rational number and decimal concepts. In T. Post (Ed.), *Teaching mathematics in grades K-8: Research-based methods* (2nd ed., pp. 201–248). Boston: Allyn and Bacon.
- Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality 1. *Bulletin de Methodologie Sociologique*, 66(1), 60–75.
- Brown, A. L. (1992). Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Bull, S. (2004). Supporting learning with open learner models. *Planning*, 29(14), 1.
- Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). Effectiveness of Reading and Mathematics Software Products: Findings From Two Student Cohorts. NCEE 2009-4041. National Center for Education Evaluation and Regional Assistance
- Clancey, W. J. (1987). Knowledge-based tutoring: the GUIDON program. MIT press
- Collins, A. (1992). *Toward a design science of education* (pp. 15–22). Heidelberg: Springer Berlin.
- Corbett, A., McLaughlin, M., & Scarpinato, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modeling and User-Adapted Interaction*, 10(2–3), 81–108.
- Cuban, L. (1986). Teachers and machines: The classroom use of technology since 1920. Teachers College Press
- Dewey, J. (1938). *Experience & Education*. New York: Kappa Delta.
- Doignon, J.-P. & Falmagne, J.-C. (1999). Knowledge Spaces. Springer-Verlag
- Du Boulay, B., & Luckin, R. (2001). Modelling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence in Education*, 12(3), 235–256.
- Evens, M. W., Chang, R. C., Lee, Y. H., Shim, L. S., Woo, C. W., Zhang, Y., & Rovick, A. A. (1997, March). CIRCSIM-Tutor: An intelligent tutoring system using natural language dialogue. In *Proceedings of the fifth conference on Applied natural language processing: Descriptions of system demonstrations and videos* (pp. 13–14). Association for Computational Linguistics.
- Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Siegler, R., & Changas, P. (2013). Improving at-risk learners’ understanding of fractions. *Journal of Educational Psychology*, 105(3), 683.

- Garellick, B. (2006). Miracle math: a successful program from Singapore tests the limits of school reform in the suburbs. *Education Next*, 6(4), 38–45.
- Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: a simulation of a human tutor. *Cognitive Systems Research*, 1(1), 35–51.
- Graesser, A. C., Jackson, G. T., Mathews, E. C., Mitchell, H. H., Olney, A., Ventura, M., & Tutoring Research Group. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society (pp. 1–6).
- Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., & Robyn, A. (2007). Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states. Rand Corporation
- Hattie, J. & Gan, M. (2011). Instruction based on feedback. Handbook of research on learning and instruction, 249–271.
- Heffernan, N. T., & Koedinger, K. R. (2002). *An intelligent tutoring system incorporating a model of an experienced human tutor* (Intelligent Tutoring Systems, pp. 596–608). Heidelberg: Springer Berlin.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- International Mathematical Olympiad (2013). Retrieved from: <http://www.imo-official.org/results.aspx>
- Jackson, G. T., Boonthum, C., & McNamara, D. S. (2009). iSTART-ME: Situating extended learning within a game-based environment. In Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education (pp. 59–68).
- Kalyuga, S. (2010). Schema acquisition and sources of cognitive load. *Cognitive load theory*, 48–64
- Karp, A. & Vogeli, B. R. (Eds.). (2010). Russian mathematics education: history and world significance (Vol. 4). World Scientific.
- Karp, A., & Zvavich, L. (2011). On the mathematics lesson. *Russian Mathematics Education: Programs and Practices*, 5, 1.
- Kay, J., Halin, Z., Ottomann, T., & Razak, Z. (1997). Learner know thyself: student models to give learner control and responsibility. In Proceedings of International Conference on Computers in Education (pp. 17–24).
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education (IJAIED)*, 8, 30–43.
- Koedinger, K. & Corbett, A. Cognitive Tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*, Cambridge University Press. 2006.
- Lee, S. W., Kelly, K. E., & Nyre, J. E. (1999). Preliminary report on the relation of students' on-task behavior with completion of school work. *Psychological Reports*, 84(1), 267–272.
- Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997, March). The persona effect: affective impact of animated pedagogical agents. In Proceedings of the ACM SIGCHI conference on human factors in computing systems (pp. 359–366). ACM.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah: Lawrence Erlbaum Associates.
- Major, N., Ainsworth, S., & Wood, D. (1997). REDEEM: Exploiting symbiosis between psychology and authoring environments. *International Journal of Artificial Intelligence in Education*, 8(3/4), 317–340.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1), 14.
- Mayer, R., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. *Cognitive load theory*, 132.
- McNeil, N. M., Grandau, L., Knuth, E. J., Alibali, M. W., Stephens, A. C., Hattikudur, S., & Krill, D. E. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction*, 24(3), 367–385.
- Middley, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes toward mathematics before and after the transition to junior high school. *Child development*, 981–992
- Milgram, R. J. (2005). The Mathematics Pre-Service Teachers Need to Know. Retrieved from: <ftp://math.stanford.edu/pub/papers/milgram/FIE-book.pdf>.
- National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. US Department of Education
- Ocuppaugh, J., Baker, R. S., Gaudino, S., Labrum, M. J., & Dezendorf, T. (2013). Field Observations of Engagement in Reasoning Mind. To appear in Proceedings of the 16th International Conference on Artificial Intelligence and Education.

- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2013). Effectiveness of Cognitive Tutor Algebra I at Scale. *Educational Evaluation and Policy Analysis*, 0162373713507480
- Pardos, Z. A., Dailey, M. D., & Heffernan, N. T. (2011). Learning what works in ITS from non-traditional randomized controlled trial data. *International Journal of Artificial Intelligence in Education*, 21(1), 47–63.
- Polyakova, T. (2010). *Mathematics education in Russia before the 1917 revolution* (pp. 1–42). History and World Significance. In Russian Mathematics Education.
- Provornik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). Highlights from TIMSS 2011: Mathematics and Science Achievement of US Fourth-and Eighth-Grade Students in an International Context. NCES 2013-009. National Center for Education Statistics
- Rau, M., Alevan, V., Rummel, N., & Rohrbach, S. (2012). Sense making alone doesn't do it: fluency matters too! its support for robust learning with multiple representations. In S. Cerri, W. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (Vol. 7315, pp. 174–184)). Heidelberg: Springer Berlin.
- Rau, M. A., Scheines, R., Alevan, V., & Rummel, N. (in press). Does Conceptual Understanding Enhance Acquisition of Fluency Or Vice Versa? Searching for Models to Investigate Mediators. To appear in the Proceedings of the Learning Analytics and Knowledge 2013 Conference (LAK 2013).
- Ravitch, D. (2001). *Left back: A century of battles over school reform*. Simon and Schuster
- Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., ... & Rasmussen, K. P. (2005). The Assistent project: Blending assessment and assisting. In Proceedings of the 12th Annual Conference on Artificial Intelligence in Education (pp. 555–562).
- Razzaq, L., Patvarczki, J., Almeida, S. F., Vartak, M., Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). The assistant builder: supporting the life cycle of tutoring system content creation. *Learning Technologies, IEEE Transactions on*, 2(2), 157–166.
- Renkl, A. (2011). Instruction based on examples. *Handbook of research on learning and instruction*, 272–295
- Ritter, S., Kulikowich, J., Lei, P., McGuire, C. L., & Morgan, P. (2007). What evidence matters? a randomized field trial of cognitive tutor algebra I. *Frontiers in Artificial Intelligence and Applications*, 162, 13.
- Rittle-Johnson, B. & Koedinger, K.R. (2009). It's not a one-way street: Iterating between conceptual and procedural lessons can improve mathematics knowledge. *British Journal of Educational Psychology*
- Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Alevan, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review*, 22(4), 379–392.
- Schmidt, W., Houang, R., & Cogan, L. (2002). A coherent curriculum. *American Educator*, 26(2), 1–17.
- Schmittau, J., & Morris, A. (2004). The development of algebra in the elementary mathematics curriculum of V.V. Davydov. *The Mathematics Educator*, 8(1), 60–87.
- Schoenfeld, A. H. (2004). The math wars. *Educational policy*, 18(1), 253–286.
- Shulman, L. S. (1987). Knowledge and teaching: foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Stefimova, N. (2010). The preparation of mathematics teachers in Russia: past and present. *Russian Mathematics Education: History and World Significance*, 4, 279.
- Stevenson, H. (1994). *Learning Gap: Why Our Schools Are Failing And What We Can Learn From Japanese And Chinese Education*. Simon & Schuster
- Stigler, J. W. & Hiebert, J. (2009). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. Free press
- Tatto, M. T., Schwille, J. S., Senk, S., Ingvarson, L. C., Peck, R., & Rowley, G. L. (2009). Teacher education and development study in mathematics (TEDS-M): Conceptual framework. IEA
- Toom, A. (1993). A Russian teacher in America. *Journal of Mathematical Behavior*, 12(2), 117–39.
- Uro, G. & Barrio, A. (2013). English Language Learners in America's Great City Schools: Demographics, Achievement, and Staffing. Council of the Great City Schools
- Valcarce, M. (2012). Russian Mathematical Pedagogy in Reasoning Mind. Retrieved from: http://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1029&context=hmc_theses
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The andes physics tutoring system: lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3), 147–204.
- Ventura, M., Shute, V., & Kim, Y. J. (2013). *Assessment and learning of qualitative physics in Newton's playground* (Artificial Intelligence in Education, pp. 579–582). Heidelberg: Springer Berlin.
- Waxman, H.C. & Houston, W.R. (2008). An Evaluation of the 2006–2007 Reasoning Mind Program. Retrieved from: http://www.reasoningmind.org/pdf/Waxman-Houston_Report_2007.pdf

- Waxman, H.C. & Houston, W.R. (2012). Evaluation of the 2010–2011 Reasoning Mind Program in Beaumont ISD. Retrieved from: http://www.reasoningmind.org/pdf/Waxman-Houston_Report_2011.pdf
- Weber, W.A. (2003). An Evaluation of the Reasoning Mind Program at Hogg Middle School. Retrieved from: http://www.reasoningmind.org/pdf/Dr_Weber_Report_2003.pdf
- Weber, W.A. (2006). An Evaluation of the 2005–2006 Reasoning Mind Project. Retrieved from: http://www.reasoningmind.org/pdf/Dr_Weber_report_2006.pdf
- Wenger, E. (1987). Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge. Morgan Kaufmann
- Whitehurst, G. (2009). Don't Forget Curriculum. Brown Center Letters on Education. Brookings Institution
- Wu, H. (1999). Basic skills versus conceptual understanding. *American Educator*, 23(3), 14–19.
- Wu, H. (2011). Phoenix rising: bringing the common core state mathematics standards to life. *American Educator*, 35(3), 3–13.
- Zhokhov, V.I. (2007). Razrabotki urokov, normativniye i kontrol'no-metodicheskiye materialy [Lesson plans, normative and assessment-methodological materials]. Ilekxa