Perceptual learning and the technology of expertise

Studies in fraction learning and algebra*

Philip J. Kellman[a], Christine Massey[b], Zipora Roth[b], Timothy Burke[a], Joel Zucker[a], Amanda Saw[a], Katherine E. Aguero[b], and Joseph A. Wise[c]
[a]University of California, Los Angeles / [b]University of Pennsylvania / [c]New Roads School

Learning in educational settings most often emphasizes declarative and procedural knowledge. Studies of expertise, however, point to other, equally important components of learning, especially improvements produced by experience in the extraction of information: *Perceptual learning*. Here we describe research that combines principles of perceptual learning with computer technology to address persistent difficulties in mathematics learning. We report three experiments in which we developed and tested *perceptual learning modules* (PLMs) to address issues of structure extraction and fluency in relation to algebra and fractions. PLMs focus students' learning on recognizing and discriminating, or mapping key structures across different representations or transformations. Results showed significant and persisting learning gains for students using PLMs. PLM technology offers promise for addressing neglected components of learning: Pattern recognition, structural intuition, and fluency. Using PLMs as a complement to other modes of instruction may allow students to overcome chronic problems in learning.

Keywords: algebra, fluency, fractions, learning technology, mathematics instruction, mathematics learning, pattern recognition, perception, perceptual learning, perceptual learning module (PLM)

1. Introduction

What does it mean to learn? To understand? To have expertise in some domain? Although approaches to mathematics teaching and learning vary widely, virtually all current approaches emphasize some combination of declarative knowledge — facts, concepts, and lines of reasoning that can be explicitly verbalized — and procedural knowledge — sequences of specified steps that can be enacted. Verbalizable knowledge may include memorized facts or co-constructed explanations, and procedures may be invented by learners or taught by direct instruction. Regardless of the pedagogical approach used to acquire them, these kinds of learning still fit within the typology of declarative and procedural knowledge.

A primary goal of this paper is to introduce a different dimension of learning that we believe has been neglected in most instructional settings. In contrast to declarative and procedural learning, we focus on *perceptual learning*, which refers to experience-based improvements in the learner's ability to extract structural patterns and relationships from inputs in the environment.[1] Rapid, automatic pick-up of important patterns and relationships –including relations that are quite abstract — characterizes experts in many domains of human expertise. Experts tend to see at a glance what is relevant to a problem and to ignore what is not. They tend to pick up relations that are invisible to novices and to extract information with low attentional load. From the standpoint of conventional instruction, the expert's fluency is mysterious — attainable only by long experience or "seasoning". Yet the passage of time is not a satisfactory explanatory mechanism for cognitive change.

We believe that persistent problems in mathematics learning, including difficulties in retention, failure to transfer, lack of fluency, and poor understanding of the conditions of application of knowledge, might be improved by systematically introducing perceptual learning interventions. In this article we consider the hypotheses that (1) some perennial difficulties in learning and instruction derive from an incomplete model of learning, specifically a neglect of *perceptual learning*, and (2) perceptual learning can be directly engaged, and accelerated, through appropriate instructional technology.

1.1 Perceptual learning

Perceptual learning (Gibson 1969) refers to experience-induced improvements in the pick-up of information. Unlike most computer-based sensor systems, which pick up information using unchanging routines,[2] humans have an astonishing ability to change their information extraction to optimize particular tasks. Although seldom mentioned in discussions of instruction or learning technology, perceptual learning underlies many, if not most, of the profound differences between experts and novices in any domain — differences such as rapid selection of task-relevant information, pick-up of higher-order relations and invariance, and effective classification.

Perceptual learning (PL) actually involves several kinds of improvements in information processing (Gibson 1969; Goldstone 1998). Kellman (2002) has argued that these may be broadly categorized in terms of *discovery* effects and *fluency* effects. Table 1 shows some of these effects and categorizes them according to this dichotomy. Discovery effects refer to learners finding the information that is most relevant to a task. One well-known discovery effect is increased attentional selectivity. With practice on a given task, learners come to pick up the relevant information for classifications while ignoring irrelevant variation (Gibson 1969; Petrov, Dosher, and Lu 2005). Practice also leads learners to discover invariant or characteristic relations that are not initially evident (cf. Chase and Simon 1973) and to form and process higher level units (Goldstone 2000; for reviews, see Gibson 1969; Goldstone 1998; Kellman 2002). These discovery processes, while seldom addressed explicitly in school learning, are pervasive, natural forms of learning. When a child learns what a dog, toy, or truck is, this kind of learning is at work. From a number of instances, the child extracts relevant features and relations. These allow later recognition of previously seen instances, but more important, even a very young child quickly becomes able to categorize *new* instances. Such success implies that the learner has discovered the relevant characteristics or relations that determine the classification. As each new instance will differ from previous ones, learning also includes the ignoring of irrelevant differences.

Fluency effects refer to changes in the efficiency of information extraction rather than discovery of the relevant information. Practice in classifying leads to fluent and ultimately automatic processing (Schneider and Shiffrin 1977), where automaticity in PL is defined as the ability to pick up information with little or no sensitivity to attentional load. As a consequence, perceptual expertise may lead to more parallel processing and faster pickup of information.

Table 1. Some characteristics of Expert and Novice information extraction. *Discovery* effects involve learning and selectively extracting features or relations that are relevant to a task or classification. *Fluency* effects involve learning to extract relevant information faster and with lower attentional load. (See text.)

	Novice	Expert
Discovery effects		
Selectivity:	Attention to irrelevant and relevant information	Selective pickup of relevant information / Filtering
Units:	Simple features	“Chunks” / Higher-order relations
Fluency effects		
Search type:	Serial processing	More parallel processing
Attentional load:	High	Low
Speed:	Slow	Fast

The distinction between discovery and fluency effects is not razor sharp. For example, becoming selective in the use of information (a discovery effect) surely increases efficiency and improves speed (fluency effects). Nonetheless, clear cases of each category are evident. Experimentally, one might expect to see effects of discovery in pure accuracy measures (without time constraints), whereas fluency changes may be more evident in speed (or speed/accuracy relations when time constraints are present).

PL should not be considered a detached aspect of learning. Rather, it is intertwined with, in fact *presupposed by,* declarative and procedural knowledge. To be useful, both facts and procedures need to be deployed in relevant situations. Relevance depends on classifying the situation. In a geometry problem, one might recall the theorem specifying that a triangle having two equal sides must also have two equal angles. Whether this recollection is immediately useful or merely distracting, however, depends entirely on classifying the situation at hand. Classifying depends on picking up information about the structure of a problem or situation. The abilities to classify, discriminate, recognize patterns, and notice invariance in new instances are exactly the abilities that improve in task-specific fashion via PL (Gibson 1969; Kellman 2002). Applying procedures also depends on pattern recognition. For example, some leading approaches to computer-based learning (e.g., Anderson et al. 1992; Anderson, Corbett, Koedinger, and Pelletier 1995) have emphasized the analysis of learning content into sets of particular procedures ("productions," in a production-system approach). Instruction then consists of teaching these productions that make up the "cognitive model" for the task. Implicit in these approaches is the need for the learner to come to *recognize* the situations in which particular procedures apply. This task is not directly instructed in most applications, yet it is a crucial complement to the learning of procedures. When concrete instances reoccur, classifying or recognizing can be merely a matter of specific memory, but in real-world tasks, this is seldom the case. More commonly, problem-solving situations vary in many particulars but possess underlying structures that determine which procedures can be fruitfully applied. For the learner, extraction of this relevant underlying structure across variable examples is crucial. This is the role of PL, and evidence suggests such abilities change dramatically with practice and form a crucial foundation of expertise.

The PL effects listed in Table 1 are very general. They suggest that methods for addressing PL in instruction would have applications to almost any learning domain. As these characteristics of expertise are well-known, we might wonder why conventional instructional methods rarely address PL directly. Likewise, computer-based and web-based instructions mostly incorporate the traditional emphases on declarative and procedural knowledge. Substantial work has gone into making tutorial formats more realistic in computer-based learning (e.g., by incorporating realistic facial expressions in an animated tutor on screen), but technology to address PL has been missing.

In our view, the lack of focus on PL derives both from inadequate appreciation of certain dimensions of learning and from a lack of suitable techniques. We can teach, or at least present, facts and procedures, but how do we teach pattern recognition or structural intuition? Whereas some PL no doubt occurs during the consideration of examples in a lecture or in the working of homework problems, these activities are not strong methods for targeting perceptual learning.

In most learning domains, the answer for the student has been to learn the facts and procedures and then to spend time immersed in that domain. This advice applies to the student pilot who cannot judge the proper glide slope on approach to landing, the radiology resident who cannot spot the pathology in the image, the chess novice who cannot see the imminent checkmate, and the algebra student who cannot see that an expression can be simplified by using the distributive property in reverse (e.g., $(2x^2 - x + 2x - 1)$ can become $(2x - 1)(x + 1)$). The expert's magical ability to see these patterns at a glance has various names: Judgment, insight, intuition, perspicacity, and brilliance. These originate from vague sources: Experience, practice, seasoning. None of these are methods of instruction; rather, they point enigmatically to the passage of time, a range of experiences, or to an innate ability.

A special issue in teaching information extraction skills is that these often involve unconscious processing. The skilled expert who intuitively classifies a problem or grasps a complex relationship often cannot verbalize the process or content of these accomplishments. Even when the process or content can be stated, hearing the description does not give a student the expert's vision or fluency.

These limitations of instruction need not be fatal. We believe there are systematic approaches for engaging PL in instructional settings. These can be realized through a combination of PL principles and digital technology.

1.2 Research in perceptual learning

Although issues of PL have been considered off and on for more than a century (e.g., James 1890; Gibson and Gibson 1955; E. Gibson 1969), not many educational applications have flowed from this work. Since the late 1980s, there has been a resurgence of basic research in PL. Overwhelmingly, however, the contemporary focus has been on low-level, sensory aspects of information extraction (for a review, see Fahle and Poggio 2002; for a critique, see Garrigan and Kellman 2008). The reason for this focus is that sensory change can provide an important window into plasticity in the brain (e.g., Recanzone, Schreiner, and Merzenich 1993). In the most recent wave of research, there has been little effort to connect PL with issues of higher-order structure (as the Gibsons emphasized earlier) and not much integration with issues of learning and thinking in cognitive psychology.

Some efforts have been made in recent years to apply PL methods in real-world learning environments. Success has been reported in adapting auditory discrimination paradigms to address speech and language difficulties (Merzenich et al. 1996; Tallal, Merzenich, Miller, and Jenkins 1998). Tallal et al. showed that auditory discrimination training in language-learning-impaired children, using specially enhanced and extended speech signals, improved not only auditory discrimination performance but speech and language comprehension as well. Similar methods have also been applied to complex visual tasks. Kellman and Kaiser (1994) designed PL methods to study pilots' classification of aircraft attitude (e.g., climbing, turning) from primary flight displays (used by pilots to fly in instrument conditions). They found that an hour of training allowed novices to process configurations as quickly and accurately as civil aviators averaging 1000 hours of flight time. Experienced pilots also showed substantial gains, paring 60% off their response times. More recently, PL technology has begun to be applied to the learning of structure in mathematics and science domains, such as the mapping between graphs and equations, or apprehending molecular structure in chemistry (Silva and Kellman 1999; Wise et al. 2000). However, applications to middle school mathematics that we report here, specifically investigating PLMs for fraction learning and algebra, have not previously been attempted.

1.3 Elements of PLMs

The critical learning activity for PL involves classification episodes. In applications to structure in mathematics and mathematical representations, the learner may be asked to recognize or discriminate a relational structure or asked to map related structures across different representations (e.g., graphic versus numeric representations) or across transformations (e.g., algebraic transformations). In designing learning interventions based on principles of PL, we engage the learner in *large numbers of brief classification episodes* — not just one or two examples. This approach departs from common practice in mathematics classrooms in two notable ways. First, learners see many more instances of the target structures and relationships and in more contexts than would normally occur in classroom settings. There, most often, a teacher works one or two problems with the whole class, students explore a rich example in small groups, or a textbook presents a small number of worked examples in each chapter section, and students may then go on to solve problems that are similar to the model in fairly obvious ways. Often it is assumed that clear statement of relevant aspects of a problem type or procedure should be sufficient for good students to learn it. Yet, this assumption is suspect and, even when correct, refers to the declarative or procedural content with little consideration of pattern recognition skills. This is related to the second characteristic of PLMs: When PL is the instructional goal, students' time and effort is devoted to *problem recognition and classification*, rather than completing calculations and procedures to solve problems. Learning trials go quickly: A student might complete a dozen or more classification trials in the time it would take to work a problem.

Another critical feature of PL is that the learning instances must incorporate *systematic variation* across classification episodes. To allow the learner to extract invariant structure, it must appear in a variety of contexts. Irrelevant aspects of problems need to vary, so that the learner does not mistakenly correlate incidental features with the structure to be learned. The failure of conventional instruction to fulfill this requirement is responsible for many limitations in math learning, such as the familiar observation that students solve algebra problems more easily when "X" naturally ends up on the left side of the equation.

When the learning task involves discriminating among a set of target structures, particularly ones that may initially be confused with each other, learning trials should incorporate direct contrasts. Learning to discriminate among a set of items that at first look alike is a frustrating learning problem commonly faced by novices. What is more, this learning problem is often underestimated by experts who have already automatized the discriminations, without necessarily being able to articulate how they make them. Because the goal of PL is learning to pick up invariant structure across varying contexts, the learning set should include novel and varied instances. In this respect, PL differs from "drill" characterized by rote repetition. In rote repetition, the same learning items repeat over and over. In PL, particular instances ideally *never* repeat. PL thus gives the learner the ability to intuit relevant structure and relations in novel contexts, whereas rote learning does not. Motivationally, the situation also differs from rote learning. Properly arranged, the seeing of increasingly discernible structure in each new instance is exciting to the learner, as it is in natural learning situations, such as when a novice birdwatcher becomes able to recognize a new bird.

Computer-based learning technology provides a natural environment for PL interventions. It can allow learners to interact in systematic ways with large sets of examples that have the desired kinds of variability. It also allows continuous tracking of the performance of each individual learner (e.g., collecting accuracy and response time data on each trial), to evaluate progress toward mastery, and to customize the learning experience so time and effort are spent where they are most needed. These same features also make learning technology a powerful tool for conducting research on PL. Elements such as feedback, task format, learning sets, and problem sequencing can be naturally and systematically manipulated, and detailed performance data automatically collected for each user provide useful dependent measures for tracking and assessing learning.

1.4 Applying perceptual learning to high-level, symbolic, explicit tasks

We anticipate at this point a natural concern. How can PL apply to high-level, symbolic, and explicit domains such as mathematics? Perceptual aspects may be thought to apply only to low level or relatively incidental aspects of mathematics, such as the use of specific visual representations (e.g., pie slices used to teach fractions). Higher-level relations and structure are often considered non-perceptual. Moreover, mathematics is symbolic in that the relation between its representations and their meanings is often arbitrary (e.g., use of the character "4" to represent the number four). Arbitrary meanings, arguably, cannot be discovered from the pickup of information available in scenes, objects, or events — i.e., they are non-perceptual. Finally, mathematics is largely an explicit discipline. Not only is understanding important, but it is important to give reasons and proofs. If structural intuitions gotten from PL are not consciously accessible, they cannot be sufficient for mathematics.

Although these concerns are plausible, we find them to be ultimately ill-founded. With regard to the scope of perception, it is not uncommon to encounter the view that "perceptual" attributes are things like color, but relations and higher-order structure are cognitive constructs. Such ideas represent in part the long shadow of traditional empiricist theories of perception and in part a confusion of sensory properties with perceptual ones (for discussion, see J. Gibson 1966; Kellman and Arterberry 1998). We share with a number of modern theorists of perception (such as James and Eleanor Gibson, David Marr, Albert Michotte, and Gunnar Johansson) the idea that perception is not primarily about low-level sensory properties, such as color; it involves extracting information about the meaningful structures of objects, arrangements, and events. This extraction uses stimulus relations of considerable complexity. Michotte, for example, offered compelling evidence and arguments that we *perceive* causality and that perception often has an "amodal" character — i.e., it is not tied to simple, local, sensory stimulation (Michotte 1962; Michotte, Thines, and Crabbe 1964). J. Gibson (1966, 1979) was most programmatic in arguing that perception involves extraction of higher-order invariance in the service of acquiring functionally relevant information about objects, relations, and events.

Applied to mathematics, what this means is that mathematical ideas, as given in the representations we use to communicate them, have structure, and efficient processing of this structure is a crucial component of learning. There is structure in equations, for example, and also in graphs. Even fraction notation or the superscripting of a number to indicate exponentiation are structural features important to doing mathematics. If the novice fails to notice some important marking or relation, fails to select the aspects relevant to a problem, fails to map a structural feature to the correct concept, expends cognitive resources too heavily, or simply processes structure too slowly, advancement in math will be impaired.

One virtue of a higher-order, ecological view of perception is that it leads naturally to the idea that structural representations furnished by perception form the foundations of other cognitive processes (Barsalou 1999; Kellman and Arterberry 1998). Real-world learning and thinking tasks partake of both perceptual extraction of structure and symbolic thinking in seamless and cooperative fashion. Being involved with only one of these or the other may be a property of research communities but not of cognitive activities in complex tasks.

1.5 Perceptual learning and cognitive load

Some of the issues we raise regarding fluency and structure learning have been examined in the context of research on cognitive load effects in learning. Considerable evidence indicates that cognitive load is an important determinant of learning and performance in various domains (Chandler and Sweller 1991), including mathematics learning. In problem solving contexts, manipulations as straightforward as combining, rather than separating, textual information and diagrams can make an appreciable difference in outcomes (Sweller, Chandler, Tierney, and Cooper 1990). Presumably, such effects indicate that the demands of extracting information or processing relations in a learning or problem solving situation may exceed limits in attentional or working memory capacity.

Most efforts to ameliorate cognitive load limits in instruction have focused on altering instructional materials. In learning or problem solving, performance may be improved by combining graphics and text (Chandler and Sweller 1991), using visual and auditory channels in ways that expand capacity (Mayer and Moreno 1998), or presenting passively viewed worked examples (Paas and van Merrienboer 1994; Sweller, Chandler, Tierney, and Cooper 1990). The value of such interventions has been clearly demonstrated. Our approach, however, suggests another avenue for escaping cognitive load limits: Changing the learner. It has long been known that practice in information extraction leads to faster grasp of structure (Chase and Simon 1974) with lower cognitive load (Shiffrin and Schneider 1977), freeing up attentional capacity to organize the parts of a task or to allow attention to higher-order structure (Bryan and Harter 1899). PL technology has the potential to allow learners to overcome load limits and access higher level structure.

1.6 Experimental objectives

In the experiments below, we report initial attempts to apply PL concepts directly to mathematics learning in the middle and early high school years. We chose domains that are known to present difficult hurdles for many students: Reasoning and problem solving with fractional quantities, and algebra. These domains make plausible points of entry for at least two reasons. First, we suspect that a substantial part of students' learning difficulties in these areas involve structure extraction, pattern recognition, and fluency issues potentially addressable by PL interventions. Moreover, these areas are both central to the mathematics curriculum, and both form important foundations of later mathematics.

2. Experiment 1: Perceptual learning in fractions

Learning in the domain of rational numbers is complicated (e.g., Behr, Harel, Post, and Lesh 1992; Lamon 2001; Post, Behr, and Lesh 1986), and we did not take on its full scope, but rather focused on several important ideas. We selected issues that are known to be problematic for many learners and that may reveal the value of PL technology in improving learning.

Specifically, we targeted students' abilities to recognize and discriminate among structures that underlie the kinds of fraction problems commonly encountered in the upper elementary and middle school curriculum. We also addressed students' ability to map these structures across different representational formats, including word problems, fraction strips, and number sentences. In designing the instructional interventions for this study (both classroom lessons and learning software), we drew heavily on detailed analyses of the conceptual progressions involved in the development of fraction concepts and problem solving that have appeared in the research literature in recent years (e.g., Hackenberg 2007; Olive 1999, 2001; Olive and Steffe 2002; Olive and Vomvoridi 2006; Steffe 2002; Thompson 1995; Thompson and Saldanha 2003; Tzur 1999).

Consider the following two problems:

(1) 10 alley cats caught 5/7 of the mice in a neighborhood. If they caught 70 mice, how many mice were in the neighborhood?

(2) A school principal ordered computers for 10 classrooms. 5/7 of the computers came with blue mice. How many mice were blue, if there were 70 mice in all?

Both of these word problems use the same object quantities (70 mice), fraction (5/7), irrelevant number information (10), and the same order of presentation of the numeric quantities (10, 5/7, and 70). Despite these superficial similarities, the two problems have contrasting underlying structures. The first problem could be restated in a simplified way as "70 mice is 5/7 of how many mice?" while the second problem could be restated as "How many mice is 5/7 of 70 mice?" Problem (1) is what we term a "find-the-whole" problem — we know that 70 mice is 5/7 of a whole quantity and we need to use that information to figure out what that whole quantity is. Problem (2) is a "find-the-part" problem — we know that the whole quantity of mice is 70 and we need to use that information to figure out how many mice would comprise 5/7 of that whole. The structural distinction between these two problems is not transparent in the structure of the word problem, and many upper elementary and middle school students do not seem to be able reliably to extract the underlying structure and carry out a corresponding solution strategy. (Indeed, we have repeatedly observed that when students encounter a find-the-part and find-the-whole problem with similar "cover stories" in a test or classroom assignment, they will frequently complain that the teacher made an error and gave them the same problem twice.)

In Experiment 1 we targeted these issues using PL technology. A central goal of the study was to help students become fluent in recognizing and discriminating find-the-whole and find-the-part fraction problems. A second, related goal was to enable them to identify and map these abstract structures across a series of different but mathematically relevant representations. That is, whether presented with a full word problem, a simplified question, a fraction strip representation, or a set of number sentences, they should be able to identify which kind of structure it represents and connect it to the corresponding structure in the other representational formats. Our hypothesis was that fluency in structure recognition and mapping is a critical component in problem solving, and that training that focuses on achieving it will transfer to significant improvements in open-ended problem solving.

The design of this study also provided an opportunity to explore another issue related to incorporating PL approaches into the learning interventions. As described above, a critical feature in PL is exposure to a widely varying set of examples that embody the relevant structures. Naturally occurring PL situations, such as children learning categories like dog or toy or machine, indicate that PL proceeds perfectly well in complex natural environments that have not been deliberately decomposed in any particular way to facilitate the child's learning. This observation is somewhat at odds with common approaches to the design of instruction in classroom settings, in which knowledge domains are often deliberately broken down and sequenced, with simpler concepts being introduced first and then used as building blocks for more complex concepts and relationships. Also, some experimental research on PL suggests that introduction of easy cases first may facilitate learning (e.g., Ahissar and Hochstein 1997).

In research on memory and motor learning, the related issue of blocked vs. randomized learning trials has received significant attention, with findings that might seem surprising in the K-12 classroom. Schmidt and Bjork (1992), for instance, argue from a review of a number of training studies that mixing item types to be learned produces better long-term learning, as well as better ability to apply learning appropriately in a variety of circumstances. Paradoxically, mixing may actually depress performance levels during (and immediately at the end of) training, but it leads to better performance in the long run.

In this context, we considered the specific question of whether to introduce first *unit fraction* examples and problems (i.e., those involving fractions with a numerator of 1) as a simple case and then build to the more complex cases of non-unit fractions. Alternatively, unit and non-unit fractions could be introduced at the same time, so students might notice relations between them from the beginning.

With these contrasting ideas in mind — a progression from simple to complex versus mixed complexity and task variability throughout the learning period — we developed two different forms of the learning software. For one group, unit fractions were introduced first, in a series of classroom lessons and then in training sessions with PLM software that involved only unit fraction problems. Subsequently, the students in this group participated in another round of classroom instruction that introduced non-unit fractions and then worked with PLM software that intermixed unit and non-unit fractions. In a contrasting condition, students participated in classroom instruction that introduced both unit and non-unit fractions and then worked with a version of the PLM software in which both types were intermixed from the beginning.

This study also included a control group that, like the two PLM groups, participated in a full 16-lesson instructional sequence on fractions and problem solving with fractions but did not work with the PLM technology. Both the software and classroom lessons were designed with an explicit focus on structural aspects of problems involving fractions and on relating and mapping fraction concepts across different representations. The control group allowed us to ask whether deliberately introducing and developing fraction concepts and problem solving strategies from a structural point of view in teacher-led instruction is (a) effective at all in promoting learning and problem solving with fractions, (b) sufficient in itself, or (c) able to be further complemented by additional PLM training. Comparing PLM and No-PLM conditions provided an assessment of the value of the PL intervention. A pre-test, immediate post-test, and delayed post-test design allowed us to us to compare these conditions in both immediate learning gains (at the end of instruction) and also in terms of durability of learning over time.

2.1 Methods

2.1.1 *Participants*

Participants were 76 students (44 female, 32 male) who were enrolled in the 7th grade in an urban public school serving a predominantly minority low-income neighborhood. Details of their demographic profile and related information may be found in Supplementary Materials at http://www.kellmanlab.psych.ucla.edu.

2.1.2 *Design*

All students were pre-tested on a custom-designed pencil and paper assessment and then randomly assigned to conditions with the constraint that the groups have approximately equal pre-test scores. Students in all three conditions participated in a series of classroom lessons. Students in the Unit First PLM condition and the Mixed PLM condition spent a number of sessions working individually with the software. Students in the No-PLM Control group had no further learning intervention after the classroom lessons. Following the learning phase, students were given an immediate post-test. A delayed post-test was given approximately 9 weeks later. No research-related learning activities occurred between the immediate post-test and the delayed post-test.

2.1.3 *Materials*

Classroom lessons. The classroom instruction involved a series of 16 interactive lessons, each about 40 minutes long, designed and conducted by one of the authors (ZR, an experienced middle school mathematics teacher and curriculum specialist). These lessons presented a foundational introduction to fractions, with a focus on structural relationships that underlie fraction concepts. In direct instruction and in small group activities, four different representations were used to help students develop useful intuitions about and to reason quantitatively with numeric quantities involving fractions. The same representations were also used in the PLM software, so the classroom instruction also served as an orientation to the software.

After instruction on fraction concepts and representations, these were connected to problem solving situations with "find-the-whole" and "find-the-part" problems (as described above). Four kinds of representations were introduced, which were also used in the PLM software. These four representation types were termed Word Problems (WP), Simple Questions (SQ), Number Sentences (NS), and Fraction Strips (FS). Figure 1 gives an example of three of these representations for the two contrasting problem types.

The Simple Questions were open-ended questions stated in a direct, canonical form. Fraction strips were representations that summarized the information

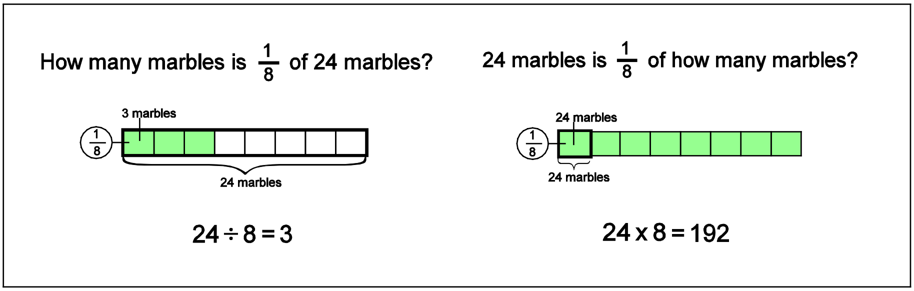


Figure 1. Examples of simple question, fraction strip, and number sentence representations for contrasting "Find-the-Part" (left) and "Find-the-Whole" (right) fraction problems. These representations were used in both the classroom instruction and PLM software in Experiment 1.

that was known in relation to the overall problem structure. The fraction strip was a continuous strip segmented according to the number of units in the fraction denominator. In the Find-the-part problem, the known quantity was the total, indicated by a labeled bracket underneath the fraction strip. In the corresponding Find-the-whole problem, the known quantity was the fractional part, indicated by a labeled bracket. Green highlighting indicated the quantity the student was trying to find. Fraction strips also included a marker that pointed to the unit fraction. The Number Sentences represented a solution strategy that could be used to find the unknown quantity.

In addition to working with the Simple Question, Fraction Strip, and Number Sentence representations, students worked on solving open-ended find-the-whole and find-the-part Word Problems, extracting a Simple Question from a Word Problem and representing the Word Problem in a Fraction Strip. Over the course of these lessons, students worked on solving a total of 10 open-ended fraction problems. The final activity in the sequence of classroom lessons involved matching all four representations to each other for both kinds of problem types. This concluding lesson also served as an orientation to the learning tasks for students in the two PLM conditions.

It is important to note that both the instructor-led classroom lessons and the learning software were created using design principles drawn from PL research: Specifically, they focused on (1) developing clear concepts of the structural relationships and patterns involved in quantities expressed as fractions, (2) the relationship between fractions and the operations of multiplication and division, and (3) recognition and mapping of target structures and patterns across representational formats. The critical differences between the classroom instruction and the PLM software were that the PLMs engaged students with a much larger and more varied set of examples, and the software-based learning experiences were designed to help students extract the target relationships on their own by interacting with them in a structured way, rather than having the learning guided and explained by a teacher. Our hypothesis was that both the classroom instruction and PLM software would advance students' learning; however, we predicted that the PLMs would enhance students' learning of structure and improve the *fluency* and *durability* of students' ability to recognize and reason with the targeted concepts.

2.1.4 *PLM software*

The PLM software presented learners with many short learning trials on which their task was to map a target structure given in one representational format to the corresponding structure in a different representational format. Learners selected from among several choices, which typically included distractor items that corresponded to common errors. Learners did not have to perform calculations or solve problems — instead the focus was on recognizing, discriminating, and mapping target structures. Figure 2 illustrates a typical learning trial.

Requiring learners to find a common structure across different representation types on each trial promotes the extraction of an abstract relational structure that cuts across superficial similarity. The choices, which were always of the same representation type, resembled each other much more than any one of them resembled the target. Thus the learner had to discriminate among stimuli with similar appearances (the choices) while mapping an abstract structure across stimuli with very different appearances (the target and its corresponding choice). The software drew on a large set of learning items so that unique items were presented on each learning trial, and memorization of the particulars of a correct answer on any

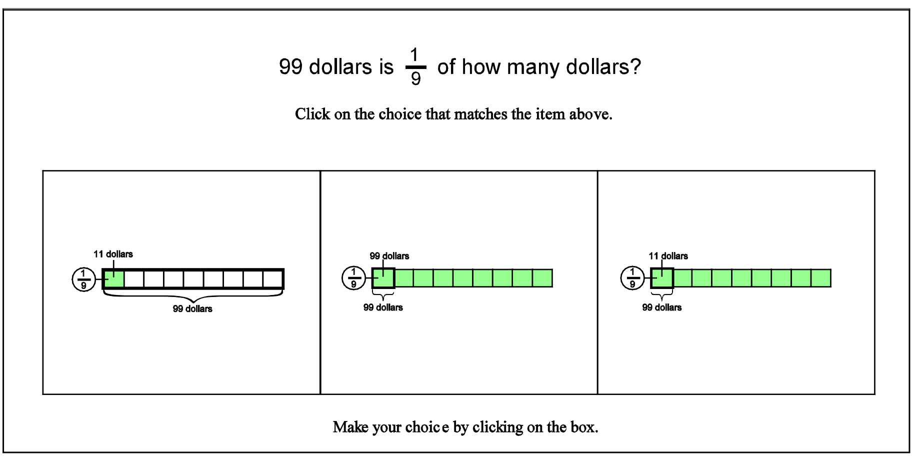


Figure 2. Sample learning trial from fractions PLM software. Learners match a target in one representational format (e.g., simple question) to the corresponding structure in another format (e.g., fraction strip). In this case, the correct choice is in the center.

given trial was not likely to help on other trials. Users received feedback on each trial as to whether they were correct or incorrect; if they were incorrect, the correct response was illustrated with a short interactive feedback sequence (described further below).

The learning set consisted of 6 categories of items, representing bidirectional pairings of each of the four representation types with each other. Learning trials contained one target representation and three choices, except for trials in which Word Problems were presented in the choice position, in which case only two choices were presented. This was done to reduce the cognitive load for learners with weaker reading skills. The program drew from a set of 112 problem families (i.e., sets of representations using the same fractions, quantities, and objects), each containing 8 potential target items and all of the related choice sets. This created a large pool of problem combinations.

The Simple Questions, because they were stated in a canonical form, had a sentence structure such that the fraction always appeared before the whole number in find-the-part problems (e.g., How many dollars is 1/5 of 20 dollars?) and vice versa for find-the-whole problems (e.g., 20 dollars is 1/5 of how many dollars?). This rigid structure may invite learners to form a rule based on the order in which the numbers appear that could guide their choice of a matching representation. To prevent such superficial rules from being useful, the Word Problems introduced the fractions and the quantities in varying orders in the same kind of problem. In addition, Word Problems included irrelevant numbers to discourage "number grabbing" strategies. These irrelevant numbers were used as distractors in corresponding incorrect choices.

Additional considerations related to constructing distractors included the use of common student errors, particularly in confusing structural relationships involved in find-the-part and find-the-whole problems. In all cases the number sentences were mathematically correct, and all fractions were fully reduced except for fractions with 100 as the denominator, which served as a bridge to thinking about percents.

The PLM software automatically created a time-stamped record of the problem presented on each trial, the student's responses, and reaction time. It also tracked the student's performance level within each category according to a set of pre-determined mastery criteria. A given category was considered to be mastered, and retired from the learning set, when the student answered 10 of the last 12 items correctly and met certain response time criteria. Time criteria were less than 90 sec per item for problems containing Word Problems and 20 seconds per item for others. As students mastered various categories, their learning effort was automatically concentrated on categories they had not yet mastered.

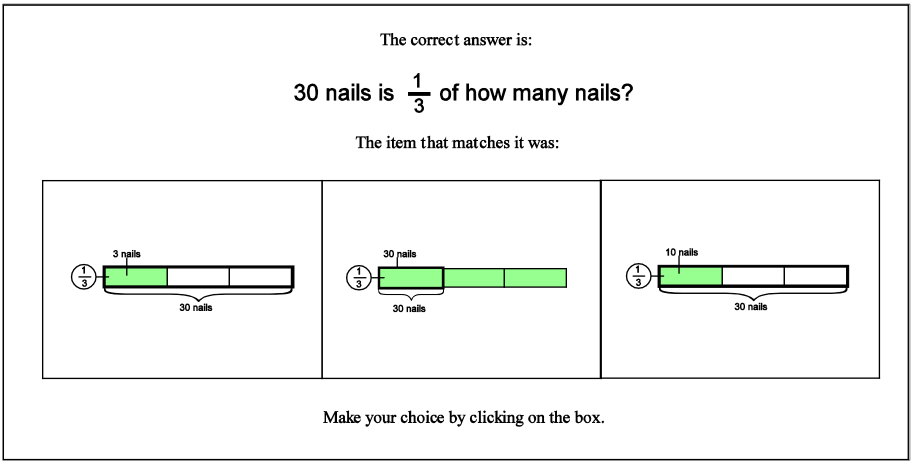


Figure 3. Active feedback screen following an incorrect response. Note that the correct response becomes the target in the active feedback on an incorrect response and the learner must match it to the original problem.

Feedback. The PLM provided students feedback on their performance in three ways: immediate feedback on accuracy, active feedback on incorrect responses, and block feedback on every twelve problems. Active feedback (see Figure 3) followed mistakes and presented the student with the correct answer again. The student was then asked to select the *question* that matches it. If the user was encoding the feedback, this selection was simple, because it had just been shown on the preceding screen. If an error occurred, the correct answer was highlighted. This active feedback was designed so that the student would have to attend to feedback information before moving on and could also gain practice on matching the representations in the opposite direction. Bi-directional practice may enhance discovery of relevant structures. Block feedback (every 12 problems) provided information on the student's accuracy and average reaction time. It also displayed a horizontal "mastery" bar that indicated (as a percentage) how close to completion the student was on the PLM. Thus, the student was able to see his or her cumulative progress.

2.1.5 *Pre-test/post-test fraction assessment*

To test for learning gains and their durability over an extended period of time, equivalent versions of a 27-item pencil-and-paper learning assessment were administered to students as a pre-test at the beginning of the study, after students had completed the learning activities for their condition, and after a delay of about two months. Items on the assessment were divided into six subscales related to different aspects of fraction knowledge and fraction problem solving. The assessment was comprised primarily of problems that did not directly resemble the kinds of problems that students worked on in either the classroom instruction or in the

PLM training and thus emphasized transfer of learning. No problem on the test was identical in structure to the learning trials included in the training. However, some — particularly the open-ended Simple Questions — were fairly close. Although students never had to solve such problems during their PLM training, they did gain considerable experience in mapping them to Number Sentences. Other problems on the assessment were less directly related to the PLM training and focused more on knowledge such as understanding unit fractions in relation to non-unit fractions and interpreting numerators and denominators in fractions. The assessment also required students to solve open-ended word problems that mixed other types of fraction problems in with find-the-whole and find-the-part problems. The subscales comprising the assessment are described in detail in the Supplementary Materials.

2.1.6 *Apparatus*

Students completed the PLM sessions on laptop PCs using the Windows operating system. The laptops were arranged on separate desks in an empty classroom at the students' school. Monitors were 13–15" in diagonal measurement.

2.1.7 *Procedure*

Classroom Instruction. Following the pre-test, students in all three conditions participated in the first round of classroom instruction involving unit fractions, which was the same for all conditions, in their regular math classes. The first round of instruction included nine lessons on unit fractions, followed by seven lessons on non-unit fractions. One of the researchers, an experienced middle school math teacher who was familiar to most of the students, designed and led the instruction with assistance from several research assistants who were available to help students as they worked on their own or in small groups. Following the first set of unit fraction lessons, students in the Unit First condition started the Unit First PLM. Simultaneously, students in the Mixed PLM and No-PLM Control conditions continued with classroom instruction that incorporated non-unit fractions. When they had completed this set of lessons, students in the Mixed PLM condition began PLM training on a version of the PLM software that intermixed unit fraction and non-unit fraction problems from the start. Students in the Unit First PLM condition completed the first phase of PLM training working only with unit fraction problems, then returned for the remaining seven classroom lessons incorporating non-unit fraction problems. They then returned to PLM training using the Mixed PLM.

PLM Sessions. Students in the Mixed and Unit-First groups were taken out of their regular classrooms for 30–40 minute sessions with the PLM software. A mini-computer lab was created using eleven laptops in an empty classroom. Students were given calculators and scrap paper but were not required to use them.

In addition to the category retirement criteria described above, the Unit First group had a group criterion in which all students had to either reach criterion within each category or complete at least 400 learning trials before all students in this group moved to Phase II. In Phase II students worked on the PLM until they either reached criterion or were stopped by the researcher due to time constraints. Students in both PLM conditions thus completed a varying number of PLM sessions, depending on their level of performance. Number of sessions ranged between 2 and 6 in Phase 1 of the Unit First PLM, 2 and 9 in Phase 2, and 2 and 13 for the Mixed PLM.

Immediate and delayed post-test administration. After reaching criterion or concluding their use of the PLM, each participant completed an immediate post-test. Students in the No-PLM Control group received their post-test following completion of instruction on non-unit fractions. Delayed post-tests were administered to all participants nine weeks later. At each administration, participants were allowed to use scrap paper and a calculator. There was no time limit, although most students completed each part of the assessment in less than thirty minutes.

2.2 Results

2.2.1 *Overall results*

The main results of Experiment 1 are shown in Figure 4. All three groups improved from pre-test to immediate post-test and delayed post-test. In the immediate post-test, the two PLM groups showed similar performance, with both outperforming the No-PLM Control Group. In the delayed post-test, however, the Mixed PLM group showed best performance, maintaining its learning gains over the 9-week interval. The No-PLM Control Group maintained its smaller learning gain after the delay. The Unit First PLM group's mean score dropped in the delayed post-test to a level lower than that of the Mixed PLM but higher than that of the control group.

These observations were confirmed by the statistical analyses. A two-way repeated measures ANOVA with Test Phase (Pre-test, Immediate Post-test, Delayed Post-test) as a within subjects factor and Condition (Unit First PLM, Mixed PLM, No-PLM Control) as a between subjects factor was performed on students' proportion correct scores on the fractions learning assessment. There was a main effect of Test Phase, $F(2,138) = 89.66$, $p < 0.001$. There was no reliable main effect of Condition, but there was a significant Condition by Test Phase interaction, $F(4,138) = 5.396$, $p < 0.001$, indicating different learning effects across conditions.

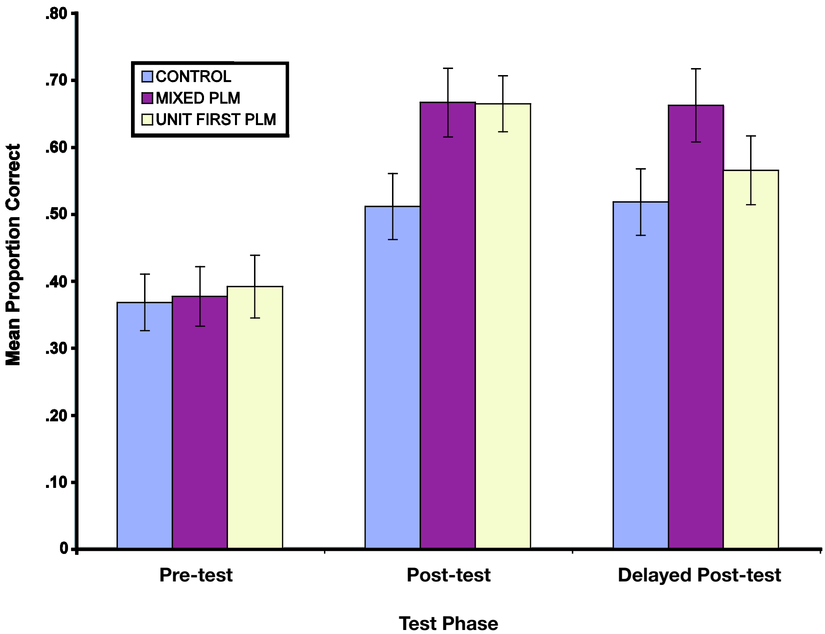


Figure 4. Mean accuracy by condition and time of test on the fraction assessment. Error bars indicate ±1 standard error of the mean.

Planned comparisons (two-tailed t-tests) were carried out to examine the condition differences in more detail. The improvement between pre-test and immediate post-test was greater in the Unit First PLM Group than in the Control Group, $t(51) = 2.60$, $p < .02$, and improvement was also greater in the Mixed PLM than in the Control Group, $t(47) = 3.07$, $p < .01$. Improvement from pre-test to immediate post-test did not differ between the Unit First PLM Group and the Mixed PLM Group, $t(48) = 0.34$, n.s.

Learning gains between the pre-test and delayed post-test did not differ reliably between the Unit First PLM and the Control groups ($t(48) = -0.528$, n.s.). However, the Mixed PLM Group showed greater improvement from pre-test to delayed post-test than both the Control Group, $t(43) = 2.86$, $p < .01$, and the Unit First PLM Group, $t(47) = 2.15$, $p < .04$.

2.2.2 *Results by subscale*

The subscales that comprise the fraction assessment provided a profile of different aspects of students' understanding. Table 2 summarizes the changes in average scores for each subscale from the pre-test to the immediate post-test and from the pre-test to the delayed post-test by condition.

Students in each condition showed substantive learning gains on all of the subscales. The largest and most durable learning gains generally favored the Mixed

Table 2. Performance on assessment subscales by condition. Pre-test columns show average proportion correct. Other columns show change from pre-test to post-test for immediate and delayed post-tests.

	Pre-test proportion correct			Change from pre-test to immediate post-test			Change from pre-test to delayed post-test		
Subscale Type	**Unit First PLM**	**Mixed PLM**	**No PLM Control**	**Unit First PLM**	**Mixed PLM**	**No PLM Control**	**Unit First PLM**	**Mixed PLM**	**No PLM Control**
Open-ended Word Problems	.30	.28	.26	+.27	+.29	+.18	+.18	+.27	+.14
Simple Find-the-part and Find-the whole Problems	.35	.33	.30	+.35	+.36	+.17	+.15	+.27	+.11
Fraction Comparisons	.64	.64	.76	+.09	+.16	−0.08	+.03	+.19	−0.13
Unit Fractions	.44	.44	.42	+.30	+.29	+.16	+.23	+.25	+.09
Find-the-Whole	.24	.32	.30	+.34	+.30	+.14	+.25	+.21	+.08
Find-the-Part	.33	.26	.24	+.28	+.34	+.15	+.12	+.31	+.17

PLM condition, which also had the highest average scores in every subscale on the delayed post-test (as can be seen by adding their change score to their pre-test score). Statistical tests showed that most of these gains were highly reliable, including 5 of 6 subscales showing robust main effects of Test Phase ($p < .001$) and 4 of 6 subscales showing a reliable interaction of Condition by Test Phase. (See Supplementary Materials.)

2.3 Discussion

On both overall assessments and individual subscales from pre-test through delayed post-test, all conditions showed significant learning gains. The Mixed PLM intervention, however, was most effective in yielding substantial learning gains (on the order of 80%) that were fully maintained more than two months later. These primary assessments were not tests of improvement in the PLM tasks but measured transfer to fundamental learning tasks, such as solving problems involving fractions and comparing fractional quantities.

These results provide clear empirical support for our motivating hypotheses. First, instruction that focuses on identifying structural patterns related to fractions, as opposed to focusing on computing solutions, is effective in leading to gains in students' ability to solve fraction problems. Although the PLM interventions required students to practice recognizing and differentiating structures, the assessments required them to solve open-ended problems. These problems were in formats differing from what students saw during the learning phase.

Second, supplementing the classroom instruction with PLM training substantially increased both overall levels of performance and the durability of learning over a two-month period. Although students in the No-PLM condition showed significant learning gains following the series of 16 classroom lessons, they did not, on average, achieve the same levels of performance as students in the PLM conditions. This suggests that classroom instruction in mathematics may aptly address some aspects of learning but not others. Declarative and procedural components need to be supplemented by learning activities in which learners practice extraction of structure and reach some level of fluency with the structures and classifications in a given domain. It suggests, further, that PLM instructional resources may be a cost-effective way to help students to attain the relevant information extraction skills and fluency. Our data indicate that students varied widely in the amount of practice needed to achieve mastery criteria. The technology introduced here constitutes an efficient way to provide varying amounts of practice to different students, as well as to monitor and certify their individual progress.

A third important finding from this experiment was that the Mixed PLM condition produced stronger learning gains than either the Unit First PLM or the No-PLM Control conditions. The Mixed PLM condition was distinctive in yielding both high levels of performance following the instructional intervention *and* long-term durability of learning. There was virtually no decrement in performance after a delay exceeding two months. The finding is noteworthy, given that the two PLM conditions were similar in many respects: Both groups experienced the same classroom lessons, and the software used by the Unit First group in Phase 2 was identical to that used by the Mixed PLM group throughout. The critical difference was that students in the Mixed PLM group saw unit and non-unit fraction problems intermixed from the beginning of their PLM training.

Why should this manipulation make such a difference? The result is consistent with earlier research indicating the value of mixed rather than blocked practice (e.g., Schmidt and Bjork 1992). In this domain, we suggest that presenting the more complete and complex learning set from the beginning allows learners to compare and contrast elements, so that their understanding of concepts such as unit and non-unit fractions and their relationship to a whole quantity is constructed in a more comprehensive and relational way from the outset. Breaking the learning apart in a simple-to-complex progression may give the learner an incomplete understanding of the elements and relationships, which must then be revised when more complexity is introduced.

3. Perceptual learning in algebra: Experiment 2

A major function of ordinary perception is to register the shapes and arrangements of objects and spatial layout in the world. Equally crucial is perceiving change, the structure of *events,* and the potential for transformation (J. Gibson 1966, 1979; Shipley and Zacks 2008). These abilities are not static: PL leads to improvements, often vast, in picking up structure, selecting relevant aspects, and becoming aware of potential for action and change (E. Gibson 1969; Goldstone 1998; Kellman 2002).

If knowledgeable scholars were asked to name specific contexts to which these descriptions of perception and PL apply, it would be a shock if any mentioned *algebra*. The idea that these concepts apply not only to ordinary perception but to higher mathematics is admittedly a novel one. Indeed, we assume we are among few if any investigators ever to suggest such a thing (but see Landy and Goldstone, in press).

On reflection, however, the idea may not be preposterous. Algebraic equations and expressions have structure, and the doing of algebra is related to the seeing of this structure. Selectivity is important: Some characteristics of algebraic representations, such as the shapes of characters, their order and arrangements, are crucial to comprehending algebra, but others, such as the size of the characters or their colors, are not. The chunking of groups of characters into meaningful entities is important in working with equations, as it is in other domains demanding perceptual expertise (e.g., Chase and Simon 1973). Efficient detection of important relations in novel exemplars — relations invisible to novices — is also key. And becoming aware of the potential for transforming equations or expressions is the bread and butter of symbolic manipulation in algebra — in simplifying an expression, classifying structure, or solving an equation.

If these aspects of learning in algebra are important, it is a matter of consequence that conventional instructional methods do little to address them directly. As in a number of complex PL domains, the expert's intuitive grasp of structure may not be available to conscious access (Chase and Simon 1973; Gibson 1969; Hoffman and Murphy 2006; Kellman 2002; Mettler and Kellman 2006). If so, the relevant pattern recognition is unlikely to be conveyed by lectures or tutorials. Even if relevant structural relations can be described, hearing the description does not turn the learner into an expert pattern recognizer. Working problems contributes to the relevant learning but perhaps not in the most systematic and efficient manner.

In two experiments, we applied PL technology to algebra learning. We were motivated by two linked hypotheses. One is that there is a learning gap in conventional instruction, such that students learn the factual and procedural aspects of algebra in their first algebra course, but are relatively impaired in terms of the seeing aspects, as might be evidenced in the speed and fluency of problem solving. The other is that PLMs providing practice in seeing structure and mapping across transformations of equations might rapidly improve these missing dimensions of instruction.

It is important to emphasize that we believe the declarative and procedural components of mathematics learning are important. Our aim is not to replace these components, but to address complementary, and neglected, learning issues, specifically those related to structure extraction and fluency. In light of this goal, in both studies we worked with students who were past mid-year in their first algebra course (Algebra I), expecting that these students would have reasonable knowledge of the facts and goals of algebra and the procedures for solving equations in one variable. We hypothesized that they might nevertheless have poor recognition skills and fluency. We predicted that a relatively short intervention, consisting of two to three days' use of a PLM with 40–45 minutes per day, might make a large and lasting difference in fluency and possibly accuracy of pattern recognition and problem solving in algebra.

Experiment 2 tested the primary hypotheses, along with tests relating the generality of problem types seen in learning to transfer. Experiment 3 was designed especially to examine endurance of learning gains, as tested after a 3-week delay, in a larger sample. Secondarily, Experiment 3 also made a first attempt to look at novel category sequencing algorithms that adapt to the individual learner.

The primary goal of Experiment 2 was to test whether principles of PL embodied in PLMs could noticeably impact algebra pattern recognition and fluency. The PLM was designed to provide practice in seeing the structure of equations and mapping across algebraic transformations. On each trial a target equation was presented, along with 4 equations shown below, labeled A through D. The participant's task was to select which of the choices could be obtained by a legal algebraic transformation of the target equation. Accuracy and speed were measured, and feedback was given.

We hypothesized that, despite offering learners no new explicit declarative or procedural information, they would show improvement from pre-test to post-test in their speed and accuracy of processing algebraic transformation problems. We also hypothesized that PLM experience might show its influence on a transfer task: Algebra problem solving. Given that solving of equations in one variable had already been taught to students, we expected that they would come to the experiment with some level of proficiency in obtaining correct answers. However, we predicted that PLM training would have a large impact on fluency, as revealed by students' speed in solving equations. A secondary objective of Experiment 2 was to look at variations in structure mapping experience. Specifically we varied the number of operations involved and direction of transformations.

3.1 Methods

3.1.1 *Participants*

Participants were 13 9th grade students and 17 8th grade students at an independent philanthropic school system in Santa Monica, California, all taking Algebra I.

3.1.2 *Apparatus*

The learning modules were tested on standard PCs using the Windows operating system in computer-equipped classrooms. Monitors were 17–21" in diagonal measurement. All assessments and the PLM were presented on computer, with participants' data being sent to a central server.

3.1.3 *Design*

The experiment was set up to assess effects of our learning technology on speed and accuracy of recognition of algebraic transformations and algebra problem solving. A pre-test was given on one day, followed by 2 days in which students worked on the PLM for 40–45 minutes per day. A post-test was administered the next day. For a subset of subjects, a delayed post-test was administered two weeks later.

3.1.4 *Algebraic transformations PLM*

In the PLM, participants on each trial selected from several choices the equation that could be obtained by a legal algebraic transformation of a target equation. An example is shown in Figure 5. Problems involved shifts of constants, variables or expressions (e.g., $(x-2)$). Accuracy and speed were measured, and feedback was given.

Participants were randomly assigned to one of four learning conditions. In Single operation conditions, participants saw problems in the learning module that always involved one operator used to transform the target equation into the correct answer (either subtract or divide). In Double operation conditions, participants saw two different operators during the learning module (on separate trials). Half of the participants in both Single and Double conditions received Unidirectional training, in that they saw problems that required transformation in only one direction, involving the shift of some constant, variable, or expression from left to right. The other half of subjects received Bidirectional training; they saw, on separate trials, either right-to-left or left-to-right transformations. Because these condition manipulations did not figure prominently in the results, we include further details in the Supplementary Materials.

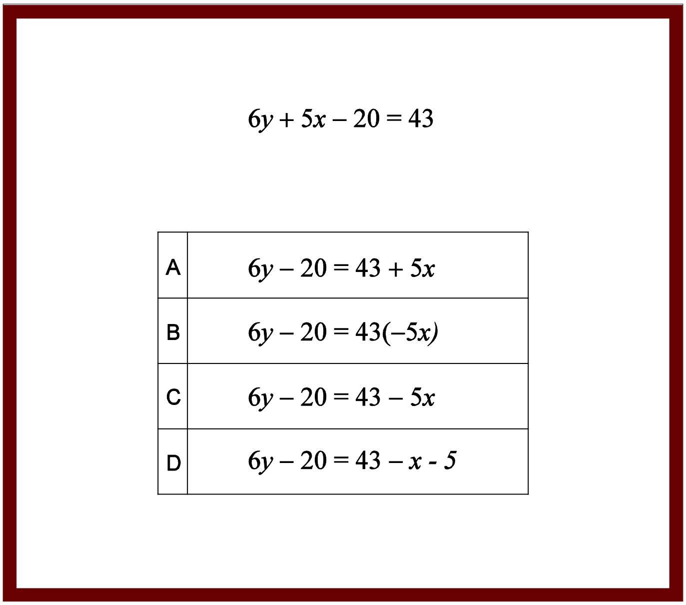


Figure 5. Sample problem from the Algebra PLM. A target equation appears at the top and the user selects which of the four choices on the bottom corresponds to a legal algebraic transformation of the target.

3.1.5 *Assessments*

All assessments were presented on the computer. Three parallel versions were constructed. Corresponding problems on separate versions varied in the specific constants, variables, or expressions appearing in each equation. Each participant saw a different version in pre-test and post-test (and, for a subset, in delayed post-test), with order counterbalanced across participants. Each version of the assessment included two sections: Recognition problems and solve problems. *Recognition* problems were similar to those in the learning module; they required the learner to select a choice that comprised a legal transformation of a target equation. In the assessments, only three choices were used — the correct answer and two distractors. *Solve* problems, requiring the participant to solve algebra equations containing one variable, were used as a transfer test. There were 16 recognition problems, 4 each involving the operators add, subtract, multiply, and divide. The left/right orientation of problems in every category was balanced. There were 17 solve problems, 8 of which were based on the single operators add, subtract, multiply, and divide. The other 9 problems were two-step problems. For example, solving $4 = 2t / 3$ might involve multiplying both sides by 3 and then dividing by 2. Participants first worked on solve-for-variable problems and then transformation problems. These problems were presented individually in random order for each participant.

3.1.6 *Procedures and stimuli*

Depending upon individual progress, participants completed the experiment (pre-test, PLM, post-test) in 2–4 sessions lasting about 45 minutes each and usually taking place on consecutive days. The interval between sessions did not exceed 2 days and the post-test was taken within 2 days of completing the PLM. All problems were presented on the computer. Scratch paper was provided for solve-for-variable problems on the pre-test and the post-test.

On each learning trial, an equation was presented at the top with an equivalent transformed equation given along with three distractor equations (making 4 choices in all). Participants were instructed to select the equivalent equation and to be accurate but respond as promptly as they could. Following either a correct or incorrect selection, portions of the original equation and the equivalent one that were relevant to the transformation were highlighted in red. If the participant chose the correct answer, a green box appeared around that choice and underneath the equations appeared the message "Correct!" and a prompt to press the spacebar to go onto the next problem. If the wrong answer was chosen, a red X crossed out the incorrect choice, a blue box surrounded the correct one, and the message "Incorrect" appeared beneath the equations. A participant timed-out on a problem if there was no response within 30 seconds. In this case, a blue box appeared around the correct answer and the message "Time is up!" appeared beneath the equations.

Following either an incorrect choice or time-out, the participant was required to interact with the feedback. A feedback screen appeared, presenting the original equation and the correct answer choice, marking in red the portions of both equations that related to the transformation. The participant was then given four choices of what operations and terms the transformation involved. If the participant made the correct choice, a green box would appear around it; otherwise, a red cross would mark out the incorrect choice and the correct choice would be highlighted with a blue box around it. The participant was then prompted to press the spacebar to continue.

After each block of 10 trials, a summary feedback screen appeared. It showed graphically the accuracy and mean reaction time for that block and up to nine preceding blocks. Participants performed a minimum of 100 learning trials and ended the learning module according to accuracy and speed criteria. These learning criteria were two blocks with accuracy ≥ 85% and an average reaction time (for correctly answered items) ≤ 8 seconds. Participants either reached learning criteria or performed a maximum of 300 learning trials before proceeding to the post-test.

For the high school participants who finished the learning module on a different day than the day of the post-test, a refresher of 30 learning trials preceded the post-test. This refresher was eliminated for the middle school participants, as it was found to be cumbersome and annoying to participants.

3.2 Results

3.2.1 *Achievement of learning criteria*

Of the 30 participants, 24 reached learning criteria. Six participants retired after 100 trials, 10 between 100 and 200 trials, and an additional eight under 300 trials. Number of operations did not markedly affect learning time but bidirectional conditions took longer than unidirectional ones (see Supplementary Materials for details).

3.2.2 *Overview of results*

The main results for performance in recognizing algebraic transformations are shown in Figures 6A and 6B, which present accuracy and response time results respectively. The PLM improved recognition accuracy from pre-test to post-test. There was a robust gain in fluency as well, with response times decreasing about 50%. Figures 7A and 7B show pre-test and post-test results for accuracy and response time on the transfer test, which involved solving open-ended algebra problems. There was little change in accuracy; students in Algebra I performed at a high level (about 80%) in the pre-test and post-test, indicating that as a group, they knew how to solve basic equations. There were, however, large changes in fluency. The data show that for solving simple equations (e.g., $3 + y = 12$) students in Algebra I take about 25 seconds per problem! Use of the PLM improved the speed of equation solving, producing an average drop in solution time of 46%. Learning ef-

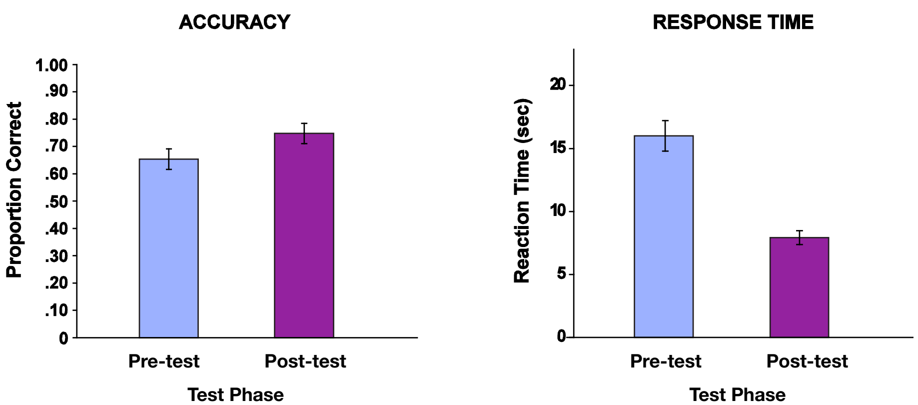


Figure 6. Students' mean accuracy and response times for recognizing algebraic transformations in the pre-test and post-test of Experiment 2. Error bars indicate ±1 standard error of the mean.

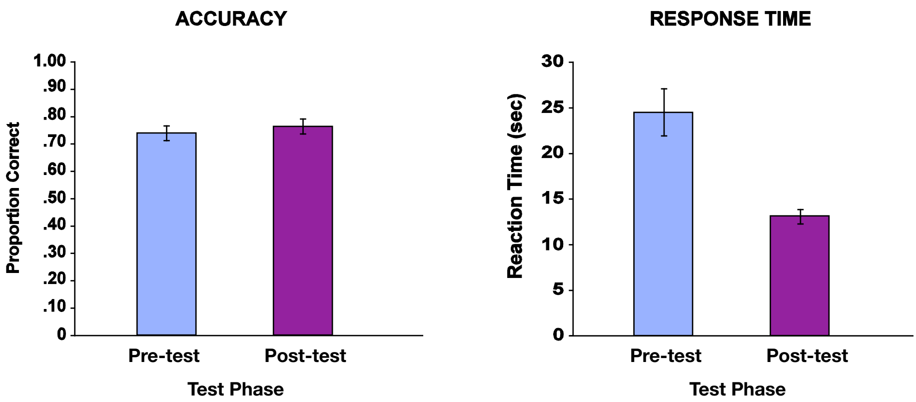


Figure 7. Students' mean accuracy and response times on transfer problems involving solving algebra equations in the pre-test and post-test of Experiment 2. Error bars indicate ±1 standard error of the mean.

fects were for the most part consistent across all conditions; the variations in training conditions produced only modest differences. These findings were confirmed by the analyses, which we consider in separate sections below.

3.2.3 *Recognizing algebraic transformations: Accuracy*

The recognition problems presented in the pre-test and post-tests resembled those presented during the learning module. Accuracy in recognizing algebraic transformations improved in all conditions through the use of the PLM. Recognition accuracy data were analyzed in a 2 (Test Phase) by 2 (Familiar vs. Unfamiliar Problem Type) by 2 (Operators in Learning) by 2 (Transformation Directions in Learning) ANOVA, in which the first two factors were tested within subjects and the latter two between subjects. There was a significant main effect of test phase, indicating overall improvement from pre-test to post-test, $F(1,26) = 9.11$, $p < .01$. There were no other reliable main effects or interactions.

3.2.4 *Recognizing algebraic transformations: Response times*

PLM exposure led to faster performance on recognition problems in all conditions. All but one participant improved. Response times were analyzed in a 2 (Test Phase) by 2 (Familiar vs. Unfamiliar Problem Type) by 2 (Operators in Learning) by 2 (Transformation Directions in Learning) ANOVA, in which the first two factors were tested within subjects and the latter two between subjects. The analysis showed a significant main effect of test phase, $F(1,26) = 56.91$, $p < .001$, reflecting the overall improvement in speed.

3.2.5 *Solving algebra problems: Accuracy*

Accuracy of problem solving was good overall and did not change noticeably, from 74.5% overall on the pre-test to 76.9% on the post-test. Accuracy data were analyzed in a 2 (Test Phase) by 2 (Familiar vs. Unfamiliar Problem Type) by 2 (Operators in Learning) by 2 (Transformation Directions in Learning) ANOVA, in which the first two factors were tested within subjects and the latter two between subjects. The overall effect of accuracy from pre-test to post-test did not reach significance, $F(1,26) = 2.816$, $p = .105$, and there were no other reliable main effects or interactions.

3.2.6 *Solving algebra problems: Response times*

Use of the Algebraic Transformations PLM led to dramatic reductions in response time in algebra problem solving. Whereas Algebra I students after the midpoint of the course do well overall in solving simple equations, remarkably, our response time data indicate that they average 24.7 sec to do so! After the PLM, the average response for the same kinds of problems was 13.2 sec. All but two participants showed faster algebra problem solving after the PLM; most showed robust gains (median = 9.2 sec per problem). These gains appear to be lasting, as shown in a delayed post-test administered to a subset of participants (see below).

Response times were analyzed in a 2 (Test Phase) by 2 (Familiar vs. Unfamiliar Problem Type) by 2 (Operators in Learning) by 2 (Transformation Directions in Learning) ANOVA, in which the first two factors were tested within subjects and the latter two between subjects. There was a substantial main effect of test phase, $F(1,26) = 46.44$, $p < .001$, but no other reliable main effects or interactions.

3.2.7 *Delayed post-test results*

A small subset of subjects ($n = 5$) was run on a delayed post-test two weeks after working on the PLM. Figure 8 displays the results from this group. Considering the pre-test and first post-test performance, this subgroup appears reasonably representative of the complete set of participants. They vary somewhat in showing no overall accuracy gain in recognition problems. What the delayed post-tests show is that the learning gains that did occur in this group were completely preserved across a two-week delay. There is a small indication that accuracy for solving equations improved from the first to the delayed post-test. The most conspicuous result, however, is that the data suggest that PLM usage produced relatively enduring gains in fluency.

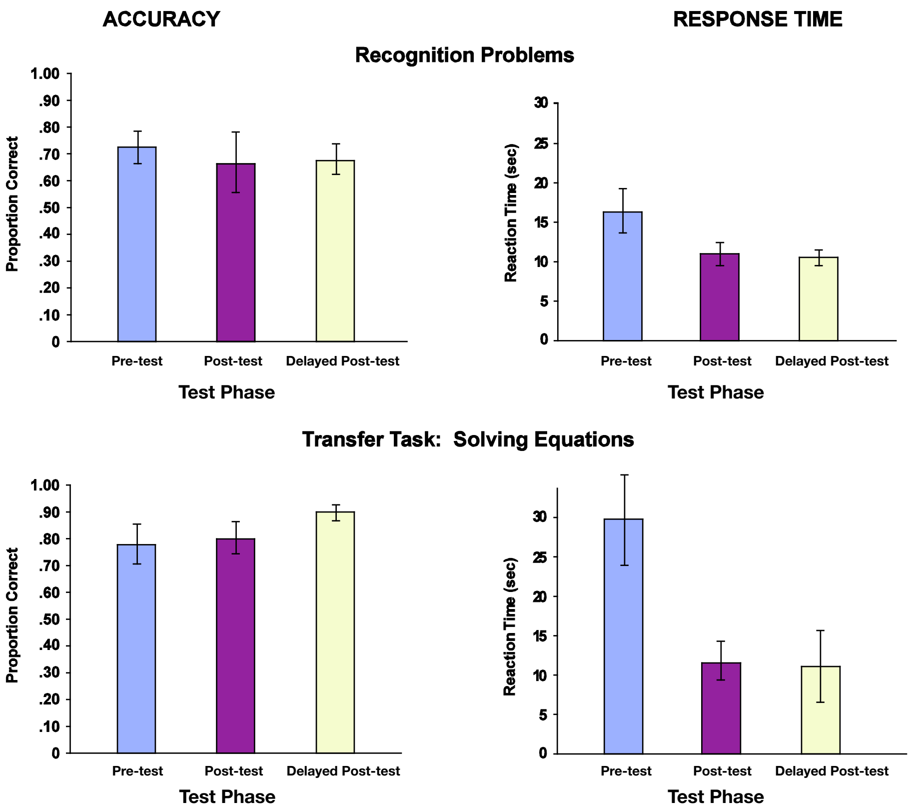


Figure 8. Delayed post-test results for Experiment 2. Pre-test, post-test, and delayed post-test data are shown for the subgroup of students who completed the delayed post-test. Top row: Mean accuracy and response time for recognizing algebraic transformations. Bottom row: Mean accuracy and response time for solving equations. Error bars indicate ±1 standard error of the mean.

3.3 Discussion

Use of the Algebra PLM produced substantial learning gains in the recognition of algebraic transformations. Both speed and accuracy improved, and almost all participants showed these gains. Practice in structure extraction and recognition of algebraic transformations also transferred to the task of solving equations. Preliminary data suggested that these effects were lasting, although only a small number of participants were tested after a delay.

These results support the two linked hypotheses that motivate these studies: That changes in information extraction (perceptual learning) are not adequately developed by conventional instruction and that technology embodying PL principles can address these missing dimensions of learning.

Although in presenting the results, we naturally focused on the latter issue — the success of the PLM in improving fluency — our data also speak to the former issue. The declarative and procedural components of learning in algebra do not directly address the perceptual extraction and pattern recognition aspects of learning. Students in this study were past mid-year in Algebra I, and pre-test data showed their high level of competence in solving a range of equations. These results for problem solving, prior to our intervention, gave a strong indication that the students knew the basic declarative and procedural requirements of basic algebra: What to do and how to do it. Yet, the pre-test data also revealed that students at this level require approximately 25 sec per problem. For an experienced adult, it is hard to fathom how someone who knows how to do algebra could take more than several seconds for problems such as $8 = x + 3$. Even for the more complicated problems in our assessment (e.g., $3x + 4 = -8$), it is hard to understand what is going on for 25 sec or more. Our surprise at these response times reflects the fact that we have acquired the experience in extracting structure and seeing transformations in this domain. Early algebra students have not. Nor do the instructional modes they normally encounter do much to facilitate these skills, at least not in the first two-thirds of the course.

Other aspects of the data point to these same crucial and neglected components of learning. Although the different learning groups (differing by number of operations seen and number of directions of transformation) did not differ much in outcomes, there was some interaction of these variables with time to complete the PLM. For example, learners who saw two directions of transformation generally took markedly longer to complete the module. The fact that direction matters (e.g., $x - 4 = 8$ vs. $8 = x - 4$) makes a crucial point about the importance of PL concepts in algebra. Mathematically, there is no important difference between $x - 4 = 8$ and $8 = x - 4$. The equal sign is symmetric, and one hopes that no mathematics teacher has ever presented these cases as different with regard to facts, concepts, or procedures. Yet, in our results, there was both a reliable main effect of mixing directions of transformation and evidence of an interaction: When learning involved more than one possible operator on different trials, it took substantially longer when bidirectional transformations were involved. These outcomes speak to the importance of the "seeing" aspects of algebra proficiency. Bidirectional transformations add to learning time because they place greater demands on the processing of structure and selection of relevant inputs in mapping across transformations.

The greater difficulty posed by bidirectional transformations also suggests connections to the well-documented finding that algebra students often do not recognize the equal sign as signifying the symmetrical relationship of equivalence; instead, they may interpret it as a signal that the preceding operation should be carried out or a marker that the answer follows (e.g., Foster 2007; Knuth et al.

2005, 2006). While interpreting the equal sign operationally rather than relationally has generally been interpreted as a *conceptual* misconception, we suggest that it may arise at least in part from PL principles and from exposure, perhaps prior to algebra, to a biased subset of examples of equations. Unfortunately, perceptual learning is not confined to the instructor's intentions: Given habitual exposure to equations in which x is on the left, students extract this directional structure and attempt to give it meaning.

The present results suggest new opportunities for combining PL technology with declarative and procedural instruction. Issues of how to optimize the technology and combine it with other modes of learning are important priorities for research. One such issue involves learning criteria. In this study, the learning phase ended when a learner achieved 85% correct or better over 20 trials, with average response time below 8.5 sec. One problem with this criterion is that, within the constraints of each condition, different problems were selected randomly. It is possible that learners could sometimes meet the criterion due to a fortuitous selection of problems. Nothing in the design guaranteed that learners had reached competence on particular types of problems. The limitations of the learning criteria may relate to one feature of the data. Although average performance on the final two blocks of the module was well above 85%, post-test performance for recognition was about 75%. Even allowing for the fact that 6 subjects did not reach criterion, these data suggest that performance was slightly lower for the wider range of problems on the post-test than at the end of learning. We explored the issue of learning criteria a bit further in Experiment 3.

4. Perceptual learning in algebra: Experiment 3

Experiment 2 indicated the promise of PLM technology in improving students' abilities to recognize algebraic transformations, and it showed that these improvements transferred to students' fluency in solving algebra problems. Experiment 3 aimed to extend these findings in several respects.

One limitation of Experiment 2 was the small size of the group receiving a delayed post-test. A number of studies have suggested that performance on an immediate post-test may not be indicative of longterm learning (Schmidt and Bjork 1992). If PL interventions are to have value as supplements to conventional mathematics instruction, it is important that they produce lasting effects. Accordingly, we tested our Algebra PLM with a larger sample of students, and we assessed performance after a 3-week delay.

Another goal of this experiment, albeit a preliminary one, was to begin to examine category sequencing in learning technology. Although this initial effort turned out to have little effect on the data, we describe it here because it was included in the design and it gives some introduction to issues of sequencing that we believe are important and which we are pursuing in other research.

Adaptive sequencing for the individual learner, we believe, is one of the greatest potential benefits of learning technology. By assessing and tracking performance through a learning module, items and concept types can be presented at the best times to boost learning for each individual. Items or categories meeting certain learning criteria can be retired from a learning set. Sequencing and retirement allow a number of laws of learning to be implemented in ways that can make learning more efficient, and more durable, for each individual. We have been working with recently patented algorithms[3] that use both the learner's accuracy and speed in short interactive trials to determine when an item or category should recur as learning proceeds. In brief, each item (or category, in category sequencing) is given a priority score that updates each trial, based on accuracy and speed of recent performance, time since last presentation, and other variables. The algorithm, tested previously on fixed learning items (e.g., basic math facts), implements several laws of learning. For example, it prohibits an item from coming up on consecutive learning trials (in order to ensure recall from long-term, rather than short-term memory). It also ensures that missed items recur fairly soon, as their learning strength was likely low. Further, to ensure durability of learning, the retention interval is stretched as learning strength increased (Landauer and Bjork 1978). To do this, response times (for correct responses) are used as a proxy for a hypothetical construct of learning strength. Specifically, the algorithm uses a function of response time to stretch the reappearance interval for that individual on that item, such that faster responses produce longer retention intervals. Earlier work on item sequencing showed it to be particularly powerful in improving learning in conjunction with retirement criteria. When an item has been answered several times accurately and faster than some criterion response time, it is removed from the learning set. The use of dynamically updating priority scores, based on trials since presentation, learner accuracy and speed for each item, allows for optimization of the efficiency of learning the entire set. For memory items (e.g., learning multiplication tables, vocabulary words, or electronic components), sequencing reduces learning time by about 50% and often improves efficiency of learning (learning gains per unit learning trial invested) by 200% or more (Kellman and Massey 2005).

Whereas much is known about laws of learning and retention for individual items, not much is known about perceptual learning of categories, where specific exemplars are novel on each trial. For example, in learning a memory item, it is not useful to have the item appear on consecutive learning trials. Because the answer is still in short-term memory when the second trial occurs, there is little gain in long-term learning strength (Karpicke and Roediger 2007; Landauer and Bjork 1978). In learning a new structural concept, however, such as whether a word in a sentence is an adverb or whether a certain molecule belongs in a certain chemical family, it may be advantageous to have consecutive learning trials relating to the same concept, at least early in learning. Thus, in the present study, we used 4 conditions: Sequenced with retirement (SR), Sequenced with retirement and blocking (SRB), Random with retirement (RR), and random with no retirement (RN). As not much is known about the optimal arrangement of category sequencing, we included blocking of trials as an additional manipulation. Whereas, in item learning, immediate reappearance of a just-tested item is known to be a poor arrangement for learning, the situation could differ in learning category structure. As each instance differs, the learner's extraction of invariance from multiple examples might be facilitated by blocking of trials. Thus, we tested cases in which sequencing operated on single presentations of each category (NB = no blocking) but also a case in which what were sequenced were 3-trial blocks from a given category (B = Blocked).

4.1 Method

The experimental methods were as in Experiment 2, except as noted below.

4.1.1 *Participants*

Participants were 56 high school students (mostly in grade 9) and 38 8th grade middle school students at the same schools as in Experiment 2, all taking Algebra I. Five students from the middle school and six students from the high school were excluded from the final data set for failing to complete all three phases of the assessments. One of the middle school subjects was removed from the study for disruptive behavior. The final sample contained 34 8th grade students, 42 9th grade students, 6 10th grade students, and 1 11th grade student. There were 40 males and 43 females in the final sample.

4.1.2 *Design*

The experiment was set up to assess effects of our learning technology on speed and accuracy of recognition of algebraic transformations and algebra problem solving. A pre-test was given on one day, followed by 2–3 days in which students worked on the PLM for 40–45 minutes per day. A post-test was administered the next day. A delayed post-test was administered to all participants three weeks later.

4.1.3 *Algebraic transformations PLM*

The PLM was similar to that in Experiment 1, with a few modifications. We reduced the number of choices for each PLM problem to 3, one correct transformed equation and two distractors. Compared to Experiment 2, problems were simplified somewhat, and all participants received all problem types in the learning phase. In addition, whereas transformations relating the target and the correct choice in Experiment 2 had all involved a shift of a term using some operator, Experiment 3 included a wider variety of items. Learning items were defined by type of operation (add, subtract, multiply, and divide) and transformation type (shift or other). Sixty percent of all problems were shift problems, in which the target "moved" from one side of an equation to another, via use of an operation. For example, for a subtract/shift problem with the target $x+5=a$, the correct choice would appear as $x=a-5$. The remaining 40% of learning items involved other kinds of transformations, such as adding a new quantity to both sides of the equation. To ensure little or no repetition of specific problems, 100 exemplars were constructed for each kind of operation.

4.1.4 *Category sequencing*

For purposes of category sequencing, problems were grouped into 8 categories. These were defined by operation (add, subtract, multiply, divide) and by whether the problems involved shifting or not.

Participants were assigned randomly to one of several presentation conditions. Conditions differed in the way categories were arranged for display. They were either selected randomly for use on each learning trial, or they were selected based on the adaptive sequencing algorithm described above, which implemented several laws of learning. Particular learning items did not repeat during the module. Adaptive sequencing involved priority scores and reappearance of *categories* of problems, with problems within categories being chosen randomly within the constraints described earlier. A separate manipulation applied sequencing either to single presentations of each category (NB = no blocking) or to 3-trial blocks from a given category (B = Blocked). A retirement feature was used to remove categories on which the learner had achieved certain learning criteria. The four conditions were then as follows: Sequenced with Retirement with single trials (SRs), Sequenced with Retirement and 3-Trial Blocks (SRb), Random, No Retirement, Blocked (RNb), and Random with Retirement, Blocked (RRb). Although there could be other combinations of these features, these choices were constrained by the limit on available test participants and our own intuitions about what might be most revealing in a first study of category sequencing.

4.1.5 *Learning criteria*

The criteria for retiring a category, when the retirement feature was used, were 5 out of the last 6 problems from that category answered correctly, with an average response time for correct responses averaging 8 sec or less. A participant finished the learning phase when these criteria were met for all 8 categories. Note that in the condition not using retirement, the learning criterion could still be implemented, although problems from "retired" categories could still be presented to the learner. For learners who did not achieve the learning criteria, the learning phase was ended on the third day, and post-test was given the next day.

4.1.6 *Assessments*

The assessments were constructed along the same lines as in the preceding experiment, with some modifications (see Supplementary Materials).

4.1.7 *Procedure*

Testing occurred at the regular class time. The first day included only a pre-test. Days 2–4 (depending on a student's achievement of learning criteria) allowed about 45 minutes per day on the PLM. PLM trials were broken into 10-trial groups, with feedback on accuracy being provided after each 10 trials. Students were encouraged to take short breaks as needed at these intervals. The post-test was given on the day following completion of the PLM. Delayed post-tests were given three weeks later.

4.2 Results

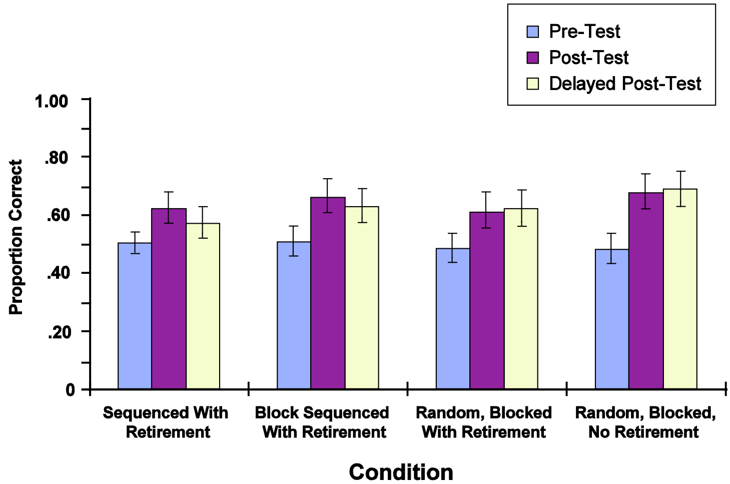
4.2.1 *Achieving learning criteria*

Out of 83 participants, 32 reached criterion. Generally, experimental effects in the group reaching retirement did not differ from those of the full group, although levels of performance were higher and learning effects somewhat clearer in the former group. For brevity, we present statistical results from the full set of subjects only, rather than two full sets of analyses.

4.2.2 *Overview of results*

Figures 9A and 9B show the accuracy and response time results for recognition problems in Experiment 3 by presentation condition. Learners showed clear improvement from pre-test to post-test in accurately mapping algebraic transformations, and these gains were mostly preserved in the delayed post-test (mean = .63). There was a robust gain in fluency as well, with response times decreasing about 29% from pre-test to post-test, and decreasing even more, about 36%, from pre-test to delayed post-test. Figures 10A and 10B show the accuracy and response

A



B

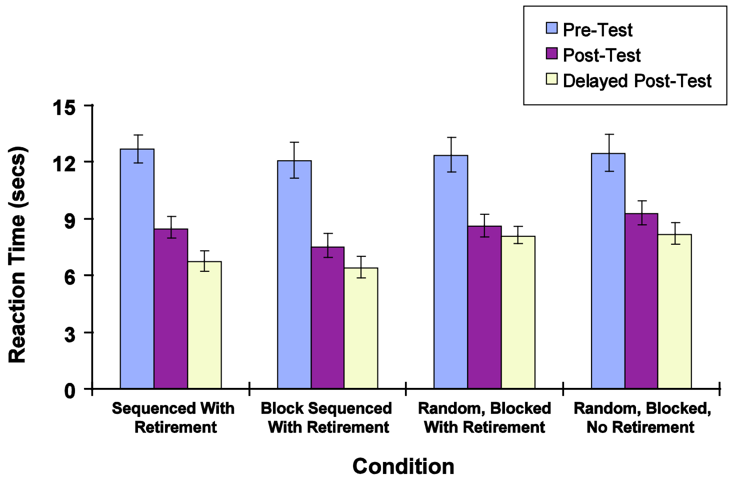


Figure 9. Students' mean accuracy (A) and response times (B) for recognizing algebraic transformations in the pre-test, post-test and delayed post-test of Experiment 3. Data are shown separately for 4 category blocking, sequencing, and retirement conditions (see text). Error bars indicate ±1 standard error of the mean.

time results for algebra problem solving before and after completing the learning module. There was no change in accuracy, but large gains in fluency: Problem solving time dropped about 32% from pre-test to post-test, and these gains were preserved in the delayed post-test. Improvements in performance in both immediate and delayed post-tests were seen in all conditions, with little effect of the sequencing, retirement, and blocking manipulations.

These findings were confirmed by the analyses, considered below.

A

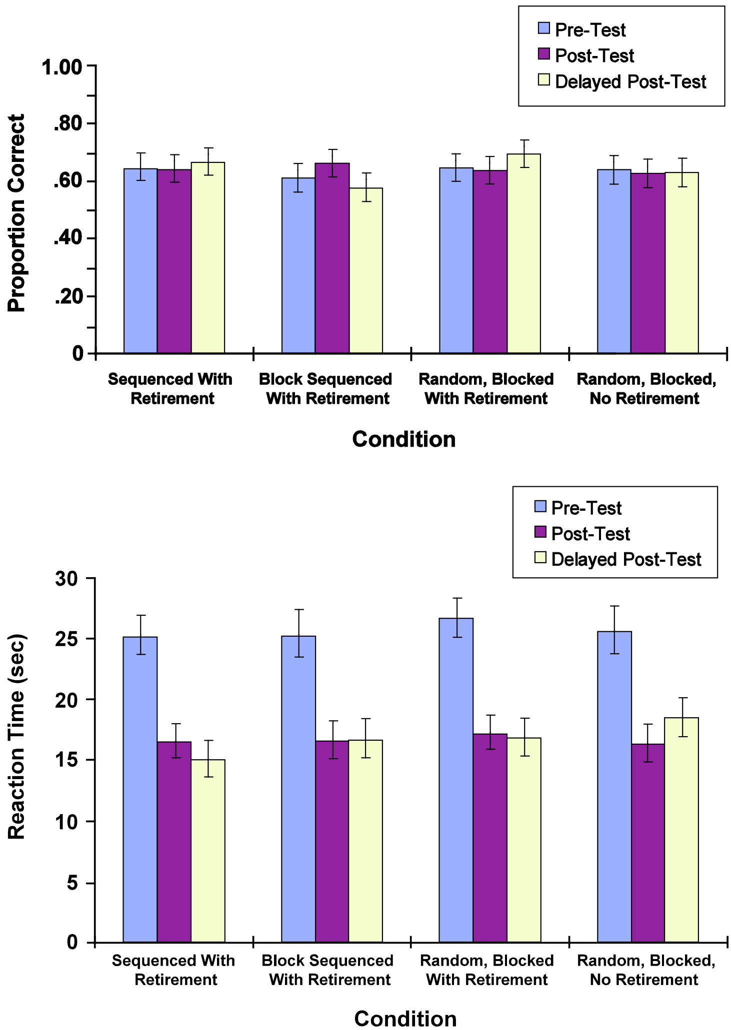


Figure 10. Students' mean accuracy (A) and response times (B) for solving equations in the pre-test, post-test and delayed post-test of Experiment 3. Data are shown separately for 4 category blocking, sequencing, and retirement conditions (see text). Error bars indicate ±1 standard error of the mean.

4.2.3 *Recognizing algebraic transformations: Accuracy*

Accuracy was analyzed via a 3 (Phase) by 4 (Condition) ANOVA with repeated measures on Phase (pre-test, post-test, delayed post-test). There was no main effect of Condition, $F(3,78) = .19$, n.s., nor any Condition by Test Phase interaction, $F(6,158) = 1.69$, n.s. The overall improvement in accuracy from pre-test to the post-tests was shown by a highly reliable main effect of Test Phase, $F(2,156) = 49.5$,

$p < .001$. Individual comparisons showed clear improvement from pre-test to immediate post-test, $t(82) = 3.82$, $p < .001$, and from pre-test to delayed post-test, $t(82) = 3.43$, $p < .001$. There was no reliable difference between performance on immediate post-test and after a three week delay, $t(82) = .35$, n.s.

4.2.4 *Recognizing algebraic transformations: Fluency*

Response times were also analyzed in a 3 (Phase) by 4 (Condition) ANOVA with repeated measures on Phase. Performance did not differ by Condition, $F(3,79) = .84$, n.s., and there was no interaction of Condition and Phase, $F(6,158) = .66$, n.s. Strong learning effects were shown by the main effect of Phase, $F(2,158) = 80.97$, $p < .001$. Individual comparisons showed clear improvement from pre-test to immediate post-test, $t(82) = 6.81$, $p < .001$ and from pre-test to delayed post-test, $t(82) = 9.06$, $p < .001$. Delayed post-test performance was also better than performance in the immediate post-test $t(82) = 2.35$, $p < .02$, suggesting that learners continued to consolidate fluency gains following the study.

4.2.5 *Solving algebra problems: Accuracy*

Transfer to problem solving accuracy did not vary as a result of the learning module, remaining level at about .65 in all phases of the study. Accuracy was analyzed via a 3 (Phase) by 4 (Condition) ANOVA with repeated measures on Phase (pre-test, post-test, delayed post-test). There was no main effect of Condition, $F(3,79) = .13$, n.s., nor any Condition by Test Phase interaction, $F(6,158) = 1.73$, n.s. There was no overall improvement in accuracy from pre-test to the post-tests (main effect of Phase), $F(2,158) = .10$, n.s.

4.2.6 *Solving algebra problems: Fluency*

Transfer to the speed of algebra problem solving showed strong effects of the PLM. Response times were analyzed in a 3 (Phase) by 4 (Condition) ANOVA with repeated measures on Phase. Learning gains did not differ by Condition, $F(3,79) = .18$, n.s., and there was no interaction of Condition and Phase, $F(6, 158) = .65$, n.s. There was a substantial main effect of Test Phase, $F(2,158) = 95.75$, $p < .001$. Individual comparisons showed clear improvement from pre-test to immediate post-test, $t(82) = 7.24$, $p < .001$ and from pre-test to delayed post-test, $t(82) = 6.92$, $p < .001$. Delayed post-test performance did not differ from that in the immediate post-test $t(82) = .08$, n.s., suggesting that learners maintained their improved skills after the experiment.

4.2.7 *Problem type analyses*

Additional analyses were carried out to investigate whether the experimental effects varied for problem types, specifically which operator was used (add, subtract, etc.), and whether the transformation involved a shift or some other change. These analyses also furnished some potentially useful baseline data about the relative difficulty of different kinds of problems for algebra students. The most notable finding was that divide problems were easier for both recognition and solve problems. (See Supplementary Materials section for details.)

4.3 Discussion

The results of Experiment 3 confirm and extend those of Experiment 2. Short interventions using PL technology improved both accuracy and speed in the recognition of algebraic transformations, and they produced conspicuous improvements in the fluency of algebra problem solving. The fluency gains in solving equations, as well as both accuracy and fluency gains for recognition problems, were fully preserved after a three-week delay.

Less informative in this experiment were our initial attempts to study categorization, sequencing, blocking, and retirement. Categorizations of problem types figured in two efforts — to begin to look at category sequencing and to be able to track particular components of learning, provide practice where it is most needed, and lead learners to meet objective criteria for each category.

Although we strongly believe in the importance and ultimate potential of these concepts, the present results make clear that their study is just beginning. There were few indications of effects of sequencing, blocking, or retirement in this study. Unfortunately, these negative findings are difficult to interpret. One reason was that fewer than half of participants met the learning criteria. Obviously, tests of manipulations involving learning criteria must be long enough to allow students to meet the criteria. Moreover, in retrospect it is not clear that the experimenters' categories (e.g., categorization of problem by operator type and by shift vs. "other" transformation) tapped different learning components for the learners. Benefits of category sequencing, blocking, and retirement hinge heavily on using categories that have validity for the learner. Finding ways to determine such categories and optimizing sequencing, blocking, and retirement schemes is a challenging but exciting priority for future research.

5. General discussion

Perceptual learning contributes enormously to expertise. It allows selective extraction of information for specific tasks, reduces required effort and attention, leads to chunking of important patterns in the input, and enables the discovery of higher-order invariance. Although these changes in information pickup can develop unsystematically through experience, attempts to address them directly in instruction have been lacking.

We hypothesized that interventions in middle school mathematics designed to foster and accelerate PL, in the form of PLM technology, might produce learning gains in pattern recognition and fluency, and that such gains might transfer to problem solving. We chose the domains of fraction learning and algebra due to their difficulty for many students and their importance in the curriculum.

The results of three experiments in two different learning domains confirm our hypotheses. In fraction learning, PLM interventions markedly improved performance on fundamental learning tasks such as solving problems involving fractions and comparing fractional quantities. Although response times were measured in the learning phase as an indicator of student progress, only accuracy in these tasks was measured in the assessments. We interpret the large learning gains across all item types in the assessments to reflect advancement of students' abilities to extract relevant relations from fractional notation and other representations, including word problems, and map them accurately across representations. The observed gains from this intervention are encouraging, especially when one considers that these students had previously had considerable exposure to fractions in their normal coursework. We found that classroom instruction combined with PLM interventions produced learning gains that exceeded and persisted beyond those gotten from classroom instruction alone. Notably, our classroom component, focusing on extracting structure and mapping structures across representations, also produced noticeable learning improvements relative to students' initial levels. This suggests that there may be value in specifically discussing relevant stuctures and structure mapping, even in conventional instructional modes. The enhanced learning produced by PLMs, however, indicates that PL technology can most directly and effectively address these aspects of learning. The results showed lasting, not transitory, changes brought about by PLMs: Learning gains survived intact over a 9-week delay.

The fraction research is special among the studies reported here in that it combined instructional techniques. This synergy underscores an important element of our approach. Although we believe that PL techniques are a sorely needed addition to most instructional contexts, they do not replace declarative and procedural components of learning. Finding the right blend of introducing facts, concepts, and procedures, along with accelerating pattern recognition and fluency through PL technology, is likely to be of maximal benefit to learning in mathematics and many other domains. The role of the teacher in introducing concepts and procedures and the role of PL technology in developing pattern recognition and fluency are complementary. Improving students' latter abilities during individual computer-based learning will allow teachers to make better use of class time. When students become fluent with basic structures and representations, their cognitive load is reduced, allowing them greater capacity to focus on new concepts or applying their knowledge. The question of how to optimize learning by combining instructional modes remains a prime question for further research.

Our experiments testing PL technology in algebra showed improvements in recognition of algebraic transformations and major gains in speed on the transfer task: Solving equations. In contrast to the fraction study, students started out with high levels of problem solving accuracy. Having recently passed the halfway point of Algebra I, they demonstrated a command of basic concepts and procedures of algebra. Students' performance illustrated vividly, however, the split between the knowing and seeing aspects of doing mathematics: In both studies, students started out requiring, on average, about 25 sec to solve simple equations. The data suggest that Algebra PLMs helped students by directly addressing pattern recognition and transformation abilities in this domain. Students practiced, not the solving of problems, but the mapping of equations onto other equations. Two to three 40-minute sessions of the Algebra PLMs improved students' accuracy and speed in recognizing algebraic transformations and produced a nearly 50% drop in the time required to solve equations. These gains proved to be lasting.

5.1 Discovery and fluency effects in PL

Earlier we distinguished discovery and fluency effects in perceptual learning. When might we see one or the other in particular PLM interventions? When learners have not yet identified what the important structures are, which parts or relations in a representation are relevant to a given task, or which features in one representation map onto those in another, one might expect that effective PLMs will produce gains in discovery, most evident in students' accuracy on assessments. Where the relevant information is known, but pattern extraction is effortful, slow and piecemeal, effective PLMs will tend to produce fluency gains.

5.2 Accuracy and speed in mathematics learning

We must also note, however, that in practical applications, such as those considered in this paper, accuracy and speed are to some degree interchangeable. As in typical psychophysical experiments, we could no doubt have changed the measured fluency differences into accuracy differences by limiting problem exposure or problem solving time. The important point for mathematics and other learning domains is not that PL discovery and fluency effects are indistinguishable. We believe these effects are different, and the present results furnish some evidence of this. The point is that both fluency and accuracy should be considered equally important in mathematics (and many other learning domains). Although this contention is hardly novel, it is unusual in educational contexts to measure response times item by item. (A test period may be time-limited, but individual problems are usually not.) The reason fluency and accuracy should be considered together is that, as students progress in mathematics, earlier learning is assumed as a foundation for new material. The student who must stop to consider the structure of fraction notation will necessarily be left behind when fractions appear in the context of a chemistry problem. And it is not just the speed at which new material appears, but its cognitive load. For many purposes, not knowing and knowing too slowly will have indistinguishable effects in impeding students' progress.

5.3 A paradox: Natural PL vs. PL technology

We claimed earlier that PL is not systematically addressed in typical instruction, but we also noted that PL is a natural, implicit learning process. The process we are trying to add into instruction and enhance through technology is the same process that allows three-year-olds to learn to classify new instances of dogs, toys, and trucks, or squares and triangles, or natural concepts of any sort. If such a process occurs naturally for three-year-olds (without lectures on the distinguishing features of dogs and cats), perhaps perceptual learning also happens in ordinary instructional settings as teachers present examples and students work problems. If so, what is the special role of systematically constructed PL technology?

There are two, related answers to this question. First, some advance in the learning and pickup of structure surely does come from seeing examples and working problems. However, this happens haphazardly, and such activities probably comprise "low doses" of PL, ultimately (perhaps years later) effective for some students but less so for others. Second, systematic targeting of PL through PLM technology can accelerate learning, through a variety of features. These include providing many more classification instances in short periods of time, systematically arranging instances to allow learning of invariance as well as learning what variation is irrelevant to a given classification, and tracking, adaptively sequencing, and retiring particular categories and classifications.

That PL technology can provide such advantages is shown in our data. In the Fractions PLM, for example, several examples and a great deal of discussion of relevant structure were presented in the control group; yet, participants in PLM groups showed greater learning gains. In the Algebra PLMs, students had seen numerous in-class and homework examples over the first 2/3 of the school year (not to mention in Pre-Algebra), but this seemed to produce little fluency, as documented in pre-test performance. Large and lasting improvements in the fluency of problem solving were produced by the PLM in only two to three class periods.

5.4 Other features of learning technology

Our discussion has emphasized novel PL techniques for mathematics instruction, but some other features of learning technology play complementary, facilitative, or in some cases, unknown roles. We required students to interact with feedback after errors, because we have learned in other work that without this feature some students may not attend much to, nor benefit much from, feedback. Although its role was not explicitly tested here (because it was always present), we highly recommend this feature for learning technology, especially when the learning activity has many short trials. Our style of reversing the question in interactive feedback may also be considered a PL manipulation, in that it again draws attention to the structure, and often, to the mapping across representations, developing flexibility in the direction of mapping. Other features, such as category sequencing, blocking, and optimal use of retirement features, were tentatively explored, but not much illuminated, by Experiment 3. These features of learning technology in a PLM format will likely prove important, but their optimal deployment will require further investigation.

We began by noting some persistent problems in mathematics learning. The present results suggest that PL technology can help address many of these issues. Theoretically, we believe PL is a neglected dimension of learning, one that requires special techniques to address systematically. The present work suggests that PLMs of relatively short duration can produce large gains in the fluency and accuracy of structure extraction, and that these gains can transfer to the fluency and accuracy of mathematics problem solving. This approach to learning technology also offers significant promise in tracking separate components of learning, continuing learning until mastery criteria are met, and incorporating response time measures to assess fluency and to include it in learning milestones. If we are correct in the assessment that most instruction does not much address PL, that pattern recognition and fluency issues arise for many students, and that these problems compound as one advances through the curriculum, then interventions made possible by principles of PL and digital technology are vitally needed. Nor do we think these issues are limited to mathematics. It is hard to think of any learning domain in which advanced performance does not rely heavily on fluent extraction of features and patterns, and on detection of invariance in changing contexts. Whether in other academic subjects, such as science or language learning, or in professional fields, such as electronics, radiology or air traffic control, the need for and promise of PL technology seem unlikely to be overestimated.

Notes

* This work was supported by the National Science Foundation under Grant REC-0231826 to PK and CM. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We gratefully acknowledge Whitney Huston, Patricia McCarrin, Eric Irwin, and Donna Perretta for their assistance and cooperation in implementing the Fractions study, and Christina Schofield and Lindsey Engle Richland for assistance in the first Algebra experiment.

1. Without further discussion, the label "perceptual" may seem overly restrictive to some. If so, the terms "structural learning" or "pattern recognition" may be used instead. In any case, what we intend should be become clear regardless of terminology, and we consider issues regarding the scope of perceptual learning below.

2. Obviously, systems that learn, either through feedback or in an unsupervised manner via the statistics of the input, would be exempt from this characterization. Such systems have been used to model human perceptual learning.

3. System and Method for Adaptive Learning, US Patent 7052277, issued May, 2006. Insight Learning Technology, Inc. holds the rights to the use of this patent. For further information, contact Insight at info@insightlearningtech.com.

References

Ahissar, M. and Hochstein, S. 1997. "Task difficulty and learning specificity". *Nature* 387: 401–406.

Anderson, J.R., Corbett, A.T., Fincham, J.M., Hoffman, D., and Pelletier, R. 1992. "General principles for an intelligent tutoring architecture". In J. Regian and V. Shute (eds), *Cognitive Approaches to Automated Instruction*. Hillsdale, NJ: Lawrence Erlbaum, 81–106.

Anderson, J.R., Corbett, A.T., Koedinger, K.R., and Pelletier, R. 1995. "Cognitive tutors: Lessons learned". *The Journal of the Learning Sciences* 4(2): 167–207.

Barsalou, L. W. 1999. "Perceptual symbol systems". *Behavioral and Brain Sciences* 22: 577–660.

Behr, M.J., Harel, G., Post, T., and Lesh, R. 1992. "Rational number, ratio, and proportion". In D. Grouws (ed), *Handbook of Research on Mathematics Teaching and Learning*. New York: Macmillan, 296–333.

Bryan, W.L. and Harter, N. 1899. "Studies on the telegraphic language. The acquisition of a hierarchy of habits". *Psychological Review* 6(4): 345–375.

Chandler, P. and Sweller, J. 1991. "Cognitive load theory and the format of instruction. *Cognition and Instruction* 8: 293–332.

Chase, W.G. and Simon, H.A. 1973. "Perception in chess". *Cognitive Psychology* 4 (1): 55–81.

Fahle, M. and Poggio, T. (eds). 2002. *Perceptual Learning*. Cambridge, MA: The MIT Press.

Foster, D. 2007. "Making meaning in algebra: Examining students' understandings and misconceptions". *Assessing Mathematical Proficiency* 53:163–176.

Garrigan, P. and Kellman, P.J. 2008. "Perceptual learning depends on perceptual constancy". *Proceedings of the National Academy of Sciences, USA*. 105(6): 2248–2253.

Gibson, E.J. 1969. *Principles of Perceptual Learning and Development*. New York: Appleton-Century-Crofts.

Gibson, J.J. 1966. *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J.J. 1979. *The Ecological Approach to Vision Perception*. Boston: Houghton Mifflin.

Gibson, J.J. and Gibson, E.J. 1955. "Perceptual learning: Differentiation or enrichment?" *Psychological Review* 62: 32–41.

Goldstone, R.L. 1998. "Perceptual learning". *Annual Review of Psychology* 49: 585–612.

Goldstone, R.L. 2000. "Unitization during category learning". *Journal of Experimental Psychology: Human Perception & Performance* 26 (1): 86–112.

Hackenberg, A.J. 2007. "Units coordination and the construction of improper fractions: A revision of the splitting hypothesis". *Journal of Mathematical Behavior* 26: 27–47.

James, W. 1890. *The Principles of Psychology (*1983 ed.*)*. Cambridge, MA: Harvard University Press.

Karpicke, J. and Roediger III, H.L. 2007. "Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(4): 704–719.

Kellman, P.J. 2002. "Perceptual learning". In R. Gallistel (ed), *Stevens' Handbook of Experimental Psychology, Vol. 3. Learning, Motivation, and Emotion, Third Edition*. New York: Wiley, 259–299.

Kellman, P.J. and Arterberry, M.E. 1998. *The Cradle of Knowledge: Development of Perception in Infancy*. Cambridge, MA: The MIT Press.

Kellman, P.J. and Kaiser, M.K. 1994. "Perceptual learning modules in flight training". *Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society*, 1183–1187.

Kellman, P.J. and Massey, C. 2005. "Background and validation research on perceptual learning and adaptive sequencing technologies". Unpublished Technical Report, Insight Learning Technology, Inc.

Knuth, E.J., Alibali, M.W., McNeil, N.M., Weinberg, A., and Stephens, A.C. 2005. "Middle school students' understanding of core algebraic concepts: Equivalence & variable". *Zentralblatt für Didaktik der Mathematik* 37(1): 68–76.

Knuth, E.J., Stephens, A.C., McNeil, N.M., and Alibali, M.W. 2006. "Does understanding the equal sign matter? Evidence from solving equations". *Journal for Research in Mathematics Education* 37(4): 297–312.

Lamon, S. 2001. "Presenting and representing: From fractions to rational numbers". In A. Cuoco and F. Curcio (eds), *The Roles of Representation in School Mathematics. 2001 Yearbook*. Reston, VA: National Council of Teachers of Mathematics, 146–165.

Landauer, T.K. and Bjork, R.A. 1978. "Optimum rehearsal patterns and name learning". In M.M. Gruneberg, P.E. Morris, and R.N. Sykes (eds), *Practical Aspects of Memory*. London: Academic Press, 625–632.

Landy, D. and Goldstone, R.L. In press. "How abstract is symbolic thought?" *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Mayer, R.E. and Moreno, R. 1998. "A split attention effect in multimedia learning: Evidence for dual processing systems in working memory". *Journal of Educational Psychology* 90: 312–320.

Merzenich, M.M., Jenkins, W.M., Johnston, P., Schreiner, C., Miller, S., and Tallal, P. 1996. "Temporal processing deficits of language-learning impaired children ameliorated by training". *Science* 271(5245): 77–81.

Mettler, E. and Kellman, P.J. 2006. "Unconscious discovery in concrete and abstract perceptual learning". *Paper presented at the Annual Meeting of the Psychonomics Society*, Houston, Texas.

Michotte, A. 1962. *The Perception of Causality*. Andover, MA: Methuen.

Michotte, A., Thines, G., and Crabbe, G. 1964. *Les Compléments Amodaux des Structures Perceptives* [The Amodal Complements of Perceptual Structures]. Louvain: Publications de l'Université de Louvain.

Olive, J. 1999. "From fractions to rational numbers of arithmetic: A reorganization hypothesis". *Mathematical Thinking and Learning* 1(4): 279–314.

Olive, J. 2001. "Connecting partitioning and iterating: A path to improper fractions". In M. van den Heuvel-Panhuizen (ed), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education* (PME-25, Vol. 4). Utrecht: Freudenthal Institute, 1–8.

Olive, J. and Steffe, L.P. 2002. "The construction of an iterative fractional scheme: The case of Joe". *Journal of Mathematical Behavior* 20: 413–437.

Olive, J. and Vomvoridi, E. 2006. "Making sense of instruction on fractions when a student lacks necessary fractional schemes: The case of Tim". *Journal of Mathematical Behavior* 25: 18–45.

Paas, G. W. C. and van Merrienboer, J. J. G. 1994. "Variability of worked examples and transfer of geometrical problem solving skills: A cognitive load approach". *Journal of Educational Psychology* 1: 122–133.

Petrov, A.A., Dosher, B.A., and Lu, Z. 2005. "The dynamics of perceptual learning: An incremental reweighting model". *Psychological Review* 112(4): 715–743.

Post, T.R., Behr, M.J., and Lesh, R. 1986. "Research-based observations about children's learning of rational number concepts". *Focus on Learning Problems in Mathematics* 8(1): 39–48.

Recanzone, G.H., Schreiner, C.E., and Merzenich, M.M. 1993. "Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys". *Journal of Neuroscience* 13(1): 87–103.

Schmidt, R.A. and Bjork R.A 1992. "New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training". *Psychological Science* 3(4): 207–217.

Schneider, W. and Shiffrin, R.M. 1977. "Controlled and automatic human information processing: I. Detection, search, and attention". *Psychological Review* 84 (1): 1–66.

Shiffrin, R.M. and Schneider, W. 1977. "Controlled and automatic information processing: II. Perceptual learning, automatic attending and a general theory". *Psychological Review* 84(2): 127–190.

Shipley, T.F. and Zacks, J.M. (eds). 2008. *Understanding Events: From Perception to Action*. NY: Oxford University Press.

Silva, A. and Kellman, P.J. 1999. "Perceptual learning in mathematics: The algebra-geometry connection". In M. Hahn and S.C. Stoness (eds), *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum, 683–689.

Steffe, L.P. 2002. "A new hypothesis concerning children's fractional knowledge". *Journal of Mathematical Behavior* 20: 267–307.

Sweller, J., Chandler, C., Tierney, P., and Cooper, M. 1990. "Cognitive load as a factor in the structuring of technical material". *Journal of Experimental Psychology: General,* 119(2): 176–192.

Tallal, P., Merzenich, M., Miller, S., and Jenkins, W. 1998. "Language learning impairment: Integrating research and remediation". *Scandinavian Journal of Psychology* 39(3): 197–199.

Thompson, P. 1995. "Notations, convention, and quantity in elementary mathematics". In J. Sowder and B. Schapelle (eds), *Providing a Foundation for Teaching Middle School Mathematics*. Albany, NY: SUNY Press, 199–221.

Thompson, P. and Saldanha, L. 2003. "Fractions and multiplicative reasoning". In J. Kilpatrick, G. Marti, and D. Schifter (eds), *Research Companion to the Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics, 95–114.

Tzur, R. 1999. "An integrated study of children's construction of improper fractions and the teacher's role in promoting the learning". J*ournal for Research in Mathematics Education* 30 (4): 390–416.

Wise, J.A., Kubose, T., Chang, N., Russell, A., and Kellman, P.J. 2000. "Perceptual learning modules in mathematics and science instruction". In D. Lemke (ed), *Proceedings of the TechEd 2000 Conference*. Amsterdam: IOS Press, 169–176.

Authors' addresses

Philip J. Kellman
UCLA Human Perception Laboratory
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1563
U.S.A.

Kellman@cognet.ucla.edu

Christine M. Massey
Institute for Research in Cognitive Science
University of Pennsylvania
3401 Walnut Street, Suite 400A
Philadelphia, PA 19104-6228
U.S.A.

massey@seas.upenn.edu

Zipora Roth
Institute for Research in Cognitive Science
University of Pennsylvania
3401 Walnut St., Suite 400A
Philadelphia, PA 19104-6228
U.S.A.

zroth@seas.upenn.edu

Timothy Burke
UCLA Human Perception Laboratory
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1563
U.S.A.

mizerai@ucla.edu

Joel Zucker
UCLA Human Perception Laboratory
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1563
U.S.A.

zuckerj@ucla.edu

Amanda T. Saw
8617 Abilene St.
Rosemead, CA 91770
U.S.A.

AmandaTSaw@aol.com

Katherine Aguero
Harlem Children's Zone – Promise Academy II
2005 Madison Ave 2nd Fl
New York, NY 10035
U.S.A.

kaguero@gmail.com

Joseph A. Wise
Center for Effective Learning
New Roads School
3131 Olympic Blvd
Santa Monica, CA 90404
U.S.A.

jwise@newroads.org

About the authors

Philip J. Kellman, PhD, is Professor and Chair of the Cognitive Area in the Department of Psychology at the University of California, Los Angeles, and Director of the UCLA Human Perception Laboratory. He is also the founder and President of Insight Learning Technology, Inc. and the author of patents in learning and display technology. His research spans basic and applied areas in visual perception, cognition, and learning.

Christine Massey, PhD, is the Director of Research and Education at the Institute for Research in Cognitive Science at the University of Pennsylvania. Her research interests connect basic research in developmental cognitive science with mathematics and science learning in educational settings.

Zipora Roth is the director of curriculum development at PENNlincs, the pre-college education group at the Institute for Research in Cognitive Science at the University of Pennsylvania. For nearly four decades she has been invested in mathematics and science learning at the elementary and middle-school levels, both as a classroom teacher and as a curriculum developer. In recent years she has focused on linking current research on learning with educational practice, creating technology-based instructional tools to support mathematics learning.

Timothy Burke is Lead Developer for the UCLA Human Perception Laboratory and Vice President of Insight Learning Technology, Inc. He is a graduate of UCLA where he received a Bachelors of Science in Cognitive Science with a Specialization in Computing and a minor in mathematics.

Joel Zucker is Director of Research for the UCLA Human Perception Laboratory and Chief Information Officer for Insight Learning Technology, Inc. He has over 20 years experience in the technology industry as an IT professional, programmer, and Head of Engineering for Lava2140, LLC. He is a graduate of UCLA and holds a Bachelors of Science in Cognitive Science.

Amanda Saw is a graduate student in cognitive psychology at the Claremont Graduate School. She is interested in learning and evaluating educational technologies. She helps to maintain and develop the Web Interface for Statistics Education website which features interactive tutorials and other statistics education resources.

Katherine Aguero graduated from the University of Pennsylvania with a B.A. in Psychology. She completed her degree of Master of Science in Education at Bank Street College of Education while teaching in Brooklyn through the Teach for America program. She is currently teaching at Promise Academy II, a charter school that is part of Harlem Children's Zone.

Joseph A. Wise is the director of The Center for Effective Learning at New Roads School. Formerly, he was a science teacher at Crossroads Upper School for Arts and Sciences for fourteen years, and served as the Science Department Chair from 1988–1996. While at Crossroads, Mr. Wise was founder and director of the W. M. Keck Math Science Institute from 1996–2001.