Topics in Cognitive Science 2 (2010) 285–305

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01053.x

Perceptual Learning Modules in Mathematics: Enhancing Students' Pattern Recognition, Structure Extraction, and Fluency

Philip J. Kellman,[a] Christine M. Massey,[b] Ji Y. Son[a]

[a]*Department of Psychology, University of California, Los Angeles*
[b]*Institute for Research in Cognitive Science, University of Pennsylvania*

Received 6 March 2009; received in revised form 18 August 2009; accepted 20 August 2009

Abstract

Learning in educational settings emphasizes declarative and procedural knowledge. Studies of expertise, however, point to other crucial components of learning, especially improvements produced by experience in the extraction of information: perceptual learning (PL). We suggest that such improvements characterize both simple sensory and complex cognitive, even symbolic, tasks through common processes of discovery and selection. We apply these ideas in the form of perceptual learning modules (PLMs) to mathematics learning. We tested three PLMs, each emphasizing different aspects of complex task performance, in middle and high school mathematics. In the MultiRep PLM, practice in matching function information across multiple representations improved students' abilities to generate correct graphs and equations from word problems. In the Algebraic Transformations PLM, practice in seeing equation structure across transformations (but not solving equations) led to dramatic improvements in the speed of equation solving. In the Linear Measurement PLM, interactive trials involving extraction of information about units and lengths produced successful transfer to novel measurement problems and fraction problem solving. Taken together, these results suggest (a) that PL techniques have the potential to address crucial, neglected dimensions of learning, including discovery and fluent processing of relations; (b) PL effects apply even to complex tasks that involve symbolic processing; and (c) appropriately designed PL technology can produce rapid and enduring advances in learning.

Keywords: Perceptual learning; Pattern recognition; Expertise; Algebra; Fluency; Learning technology; Mathematics learning; Mathematics instruction

Correspondence should be sent to Philip J. Kellman, Department of Psychology, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1563. E-mail: kellman@cognet.ucla.edu

1. Introduction

What aspects of scientific research on learning have relevance for education? In the educational literature, we find discussions of fact learning, conceptual understanding, procedure learning, constructing explanations, analogical reasoning, and problem-solving strategies. Except for an occasional mention of pattern recognition, little hints at any role for perceptual learning (PL). Likewise, studies of PL involving sensory discriminations among a small set of fixed stimuli appear to have little connection to real-world learning tasks, least of all to high-level, explicit and symbolic domains such as mathematics. Despite this apparent irrelevance or neglect, PL—improvements in information extraction as a result of practice—is one of the most important components of learning and expertise in almost any domain, including mathematics. It is also the component least addressed by conventional instruction; thus, problems of information selection and fluent extraction of structure pose major obstacles for many learners.

In this article, we describe the relevance and promise of applications of PL to high-level cognitive domains. We focus on mathematics learning, but we hope the relevance to other areas of learning will be obvious. By providing several brief examples of PL interventions in mathematics, we hope to give a useful overview of the scope, character, and possibilities of PL in instructional settings.

Applying PL to rich instructional domains is a startling and exciting endeavor. It is startling because some current conceptions of PL would preclude this kind of connection, and it is exciting because considerable evidence suggests that PL technology can address crucial and neglected dimensions of learning, offering the possibility of major advances in the effectiveness of learning in many domains.

We first summarize some characteristic improvements in human information extraction that derive from PL. Next we explain the connection between these changes in high- and low-level PL tasks, focusing on the notions of discovery and selection as unifying concepts. We then describe three examples of research on perceptual learning modules (PLMs) that apply these ideas to mathematics learning, each illustrating a separate strength of PL interventions. The current treatment is brief, and we refer the reader to other sources for more expansive discussion of some issues (Garrigan & Kellman, 2008; Kellman, 2002; Kellman & Garrigan, 2009; Kellman et al., 2008; Massey et al., in press).

2. Characteristics of expertise generated by PL

In contrast to the literature on learning, scientific descriptions of expertise are dominated by PL effects. In describing their studies of master level performance in chess, Chase and Simon (1973) wrote: ‘‘It is no mistake of language for the chess master to say that he ‘sees’ the right move; and it is for good reason that students of complex problem solving are interested in perceptual processes’’ (p. 387). Table 1 summarizes some of the well-known information processing changes that derive from PL. Kellman (2002) suggested that these can be divided into *discovery* and *fluency* effects. Discovery pertains to finding the features

Table 1
Some characteristics of novice and expert information extraction

	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		
Speed	Slow	Fast
Attentional load	High	Low
Search type	Serial processing	Some parallel processing

or relations relevant to learning some classification, whereas fluency refers to extracting information more quickly and automatically with practice. Both discovery and fluency differences between experts and novices have since been found to be crucial to expertise in a variety of domains, such as science problem solving (Bransford, Brown, & Cocking, 1999; Chi, Feltovich, & Glaser, 1981; Simon, 2001), radiology (Kundel & Nodine, 1975; Lesgold, Rubinson, Feltovich, Glaser, & Klopfer, 1988), electronics (Egan & Schwartz, 1979), and mathematics (Robinson & Hayes, 1978).

3. Scope of PL

Eleanor Gibson, who pioneered the field of PL (for a review, see Gibson, 1969), defined it as ''an increase in the ability to extract information from the environment, as a result of experience…'' (p. 3). She described a number of particular ways in which information extraction improves, including both the discovery and fluency effects noted above. Of particular interest to Gibson was ''…discovery of invariant properties which are in correspondence with physical variables'' (Gibson, 1969, p. 81). This view of PL applies directly to many real-world learning problems; although Gibson did not mention mathematics learning, her examples included chick sexing, wine tasting, map reading, X-ray interpretation, sonar interpretation, and landing an aircraft.

Much contemporary PL research differs from these earlier ideas, both in the types of learning problems studied and in theoretical context. It is common to consider PL as described by Fahle and Poggio (2002): ''…parts of the learning process that are independent from conscious forms of learning and involve structural and/or functional changes in primary sensory cortices.'' This emphasis derives from several related ideas. PL effects often show specificity to stimulus variables (such as orientation or motion direction) or observer variables (such as the eye or part of the visual field used in training). Specificity in learning (lack of transfer) has been assumed to indicate a low-level locus of learning effects, implying that mechanisms of PL are modifications, such as receptive field changes, in early sensory areas.

Confining PL to changes early in sensory systems would separate it from most real-world learning tasks involving higher-order relations and extraction of invariant patterns from varying instances. If the supporting assumptions are correct, then the type of PL being studied recently has little to do with mathematics learning. Perhaps the discovery of invariance in complex tasks and modification of receptive fields in simple discriminations have little do with each other, and the best course would be to adopt different names for these different kinds of learning.

Such a division would be unfortunate, for two reasons. First, the assumptions supporting confinement of PL to early sensory cortices and simple sensory phenomena are not sustainable. Those that are logical assumptions, such as the idea that specificity of learning implies a low-level locus, are fallacies (Mollon & Danilova, 1996). Those that are empirical hypotheses appear to be false (for more extended discussions, see Garrigan & Kellman, 2008; Kellman & Garrigan, 2009). For example, data about transfer from PL experiments have been inconsistent, and small task variations can lead to large differences in generality of learning (e.g., Liu, 1999; Sireteanu & Rettenback, 2000); single-cell recording data offer little support for receptive field modification as the explanation for PL, at least in vision (e.g., Ghose, Yang, & Maunsell, 2002). Furthermore, substantial evidence suggests that PL, even when oriented toward a low-level sensory discrimination, is guided by higher-level tasks (Ahissar & Hochstein, 2004) and attention (Seitz, Lefebvre, Watanabe, & Jolicoeur, 2005; Seitz & Watanabe, 2005), and it may in fact be required to work through constancy-based representations (Garrigan & Kellman, 2008).

Perhaps the most important rationale for considering ''high-'' and ''low-'' level tasks together in PL is that they may share common principles and mechanisms. For Gibson, a key to PL was *selection*. Recent experimental results suggest that selection is a better explanation of many low-level PL phenomena than receptive field modification (Ahissar, Laiwand, Kozminsky, & Hochstein, 1998; Petrov, Dosher, & Lu, 2005). Higher levels of processing receive inputs from below; learning processes discover which inputs are most relevant, and these inputs become more heavily weighted (Mollon & Danilova, 1996; Petrov et al., 2005). Processes of discovery and selective weighting apply to both laboratory studies of sensory discrimination and discovery of high-level invariance in ecologically important tasks.

Does specificity of learning, then, indicate a gulf between low and high-level PL? In real-world tasks, detecting structural invariance across otherwise variable cases is crucial to useful learning (e.g., detecting a wrist fracture in an unfamiliar X-ray or seeing that an algebraic transformation involves the distributive property). Yet simple laboratory PL tasks often show strong specificity of learning. We believe there is really no conflict between these observations. Specificity in learning reflects the task posed to the learner (Mollon & Danilova, 1996). If thousands of trials using only two stimuli are used, PL processes of discovery and selection may exploit local analyzers specifically relevant to those two stimuli. If accurate classification in a task requires extracting invariance and decoupling it from irrelevant, variable attributes, PL will lead to selection of relevant features or patterns that are not tied to irrelevant attributes of specific cases. For this reason, PL based on practice in extraction of structure across changing cases offers a powerful method for producing transfer of learning to novel cases. The idea that specificity in learning relates to the design and requirements of the learning task is consistent with the variability of results regarding transfer in the PL literature (e.g., Liu, 1999); with the evidence favoring selection processes over receptive field changes in PL; and with evidence and models suggesting that PL can involve information at multiple levels in a flexible manner (e.g., Ahissar & Hochstein, 2004).

4. Applications of PL to high-level learning domains

Assuming the relevance of PL to complex tasks, one might still wonder about symbolic domains such as mathematics. Even Gibson's examples—wine tasting, recognizing aircraft, or reading X-rays—seem much less ''cognitive.'' Mathematics might be thought to involve only declarative knowledge and procedures. There are inherently symbolic aspects of mathematical representations that cannot be apprehended via information ''in the stimulus.'' This is true, but it is also true that mathematical representations pose important information extraction requirements and challenges. Characteristic difficulties in mathematics learning may directly involve issues of discovery and fluency aspects of PL. A number of studies indicate the role of PL in complex cognitive domains, such as mathematics (Goldstone, Landy, & Son, 2008; Silva & Kellman, 1999), language or language-like domains (Gomez & Gerken, 1999; Reber, 1993; Reber & Allen, 1978; Saffran, Aslin, & Newport, 1996), chess (Chase & Simon, 1973), and reading (Baron, 1978; Reicher, 1969; Wheeler, 1970). Some have asserted that in general, abstract concepts have crucial perceptual foundations (Barsalou, 1999; Kosslyn & Thompson, 2000; Landy & Goldstone, 2008). Moreover, in complex cognition it is important to realize that conceptual and procedural knowledge must work together with structure extraction. Both declarative and procedural knowledge depend on pattern recognition furnished by PL. Which facts and concepts apply to a given problem? Which procedures are relevant? How do we appropriately map parts of the given information into schemas or procedures? These are fundamentally information selection and pattern recognition problems.

5. PL technology

The lack of PL techniques in instructional contexts owes not only to its neglect in learning research but also to the lack of suitable methods. The expert's pattern extraction and fluency are thought to develop separately from formal instruction, as a result of experience. Yet recent efforts suggest that there are systematic ways to accelerate the growth of perceptual expertise, in areas as diverse as aviation training (Kellman & Kaiser, 1994), medical learning (Guerlain et al., 2004), language difficulties (Merzenich et al., 1996; Tallal, Merzenich, Miller, & Jenkins, 1998), and mathematics (Kellman et al., 2008; Silva & Kellman, 1999).

In our work, we implement PL principles in PLMs. Although a full description is beyond our scope here, we mention some elements of PL interventions. Although we lack complete models of PL in complex tasks, it appears that information extraction abilities advance when the learner makes classifications and (in most cases) receives feedback. Digital technology makes possible many short trials and appropriate variation in short periods of time, allowing the potential to accelerate PL relative to less frequent or systematic exposure to structures in a domain. Unlike conventional practice in solving problems, learners in PLMs typically discriminate patterns, compare structures, make classifications, or map structure across representations.

6. Specificity of PL interventions

How do we know that a learning intervention targets PL rather than other aspects of learning? This is a complex question, and one which, in realistic instructional settings, has no absolute answer. In the work reported here, for example, a basic commitment is to use PL interventions to address core domains and known problem areas in mathematics. In doing so, less stimulus and task control are available relative to artificial materials or laboratory tasks. We have little doubt that there are some unsystematic opportunities for PL present in ordinary instruction, and, conversely, that students' declarative and procedural knowledge may interact with our PL interventions. A more global problem in making crisp distinctions is that it is likely that improved information extraction obtained through PL normally interacts strongly with other cognitive processes. Such synergy may account for the general tendency found by Fine and Jacobs (2002) for PL effects to be larger in higher-order tasks. Supporting thought and action are, after all, the functions of perception and PL. It is likely that research communities focusing on one or another of our cognitive faculties are more clearer partitioned than our use of these faculties in complex tasks.

Despite such complexities, we believe several characteristics distinguish PL interventions from conventional instruction. These characteristics involve both design of an intervention and outcomes related to characteristics of expert information extraction (Table 1).

At least three general properties are common to PL interventions:

1. *Task requiring transactions with structure*. The most basic requirement for a PL intervention is that it involves a discrimination and/or classification based on structure extracted from some representations or displays. Thus, instruction that takes the form of a verbal discussion of ancient cultures is not a promising candidate for PL. PL tasks involve practice with displays or representations in which success depends on the learner coming to attend to, discriminate, classify, or map structure. ''Use of structure'' may seem common to many aspects of instruction; a PL task, however, focuses on commonalities and variations in structure as its primary learning content. For example, in mathematics, a task requiring classification of structure can often be contrasted with a task requiring problem solving that provides a numerical answer, as in Experiments 1 and 2.
2. *Numerous classification trials with varied instances*. PL interventions involve many short trials in which the learner makes classifications and receives feedback. In

complex tasks, PL often involves ''the discovery of invariant properties which... may be buried, as it were, in a welter of impinging stimulation'' (Gibson, 1969, p. 81). Such discovery requires sufficient variation in learning instances and sufficient trials to allow relevant properties to be decoupled from irrelevant ones.

3. *Minimal emphasis on explicit instruction*. The primary task in a PL intervention does not involve verbal or written explanations of facts, concepts, or procedures. This is a major difference from conventional instruction, which is dominated by explicit description (and is certainly important). PL interventions may incorporate explicit introductions or brief discussions, but these do not comprise the central learning tasks nor are they capable of producing the results obtained with PLMs.

That an intervention impacts PL is a function of its design but also its outcomes. Some potential signatures of PL effects include the following:

1. *Generativity in structure use*. PLMs in rich learning domains are designed to improve pick-up and processing of structural invariants across variable contexts. As such, evidence of acquisition involves accurate and/or fluent classification of novel cases. Moreover, PLMs often facilitate remote transfer to different-looking problem types that involve the same underlying structure. Such transfer is a notorious problem following most conventional instructional approaches. Evidence of accurate and fluent classification of novel instances, and transfer to contexts involving different procedural requirements but common structures, provide evidence of PL.
2. *Fluency effects*. PL effects include not only selective extraction of relevant information but changes in fluency, evidenced by greater speed and automaticity, and lower effort and attentional load in information pick-up. Acquisition data within PLMs suggest that fluency in information extraction increases gradually across interactive trials. Gradual improvement is not unique to PL but does contrast with some effects of declarative instruction, in which learner may either know or not know a certain concept. Fluency effects in PL are a focus of Experiment 2.
3. *Implicit pattern recognition versus explicit knowledge*. Although PL may provide important scaffolding for explicit, verbalizable knowledge, PL itself need not involve explicit knowledge. PL changes the way a learner views a problem or representation; this idea of ''mind as pattern recognizer'' (Bereiter & Scardamalia, 1998) need not be accompanied by explicit facts, concepts, or procedures. In some domains, one might be able to demonstrate a ''double dissociation'' between PL effects and effects of conventional instruction. Whereas conventional instruction may lead to verbalizable knowledge but lagging pattern recognition and fluency, PL may produce the reverse. Such a clear division, although imaginable, may in practice be difficult to observe, because these forms of learning are normally synergistic, producing performance outcomes in which pattern recognition, facts, concepts, and procedures interact. The experiments reported here did not assess students for their abilities to generate verbal explanations, but this kind of dependent variable might be fruitfully contrasted with PL effects in future studies.

4. *Delayed testing effects*. Allegedly, one never forgets how to ride a bicycle. If true, riding a bicycle, a task that clearly involves considerable PL, differs from most declarative and procedural learning. It is not by accident that virtually all current middle and high school math textbooks begin with a long unit reviewing the previous year's content. Facts and procedures are subject to forgetting, often precipitously so. Improved facility in picking up patterns and structure in PL, like riding a bicycle, may be comparatively less subject to decay with time. This claim is conjectural, but maintenance of these skills in delayed posttests may be a hallmark of PL. Experiments 2 and 3 below examine this possibility, with a very long delay (4.5 months) in Experiment 3.

7. Experiment 1—Mapping across multiple representations: The MultiRep PLM

Mathematical representations are aimed at making concepts and relations accurate and efficient, but they pose complex decoding challenges for learners. Each representational type (e.g., a graph or an equation) has its own structural features and depicts information in particular ways. Perceptual extraction of structure from individual representations and mapping across representations present learning hurdles that are not well addressed by ordinary instruction. We developed the *Multi-Rep PLM* to help middle and high school students develop pattern recognition and structure mapping with representations of linear functions, in graphs, equations, and word problems. As in many PLMs, rather than having students solve problems for a numerical answer, we presented them with short, interactive classification tasks that facilitated fluent extraction of important features and patterns.

On each trial of the PLM, either an equation, graph, or word problem expressing a particular linear function was presented. The learner was asked to select an equivalent function among three possible choices in a different representation (e.g., if an equation was presented, the choices could be three possible graphs).[1] An example is shown in Fig. 1. There were six types of mapping trials, comprising all possible pairs of word, equation, and graphical representations given as targets and choices. Incorrect choices usually shared some values with the target display (e.g., having a common slope) but differed in some other respect (e.g., having a different y-intercept). Across problems, a variety of contexts and numerical values were used.

The rationale for the PLM was that fluent use of each representational type requires the ability to extract particular structural attributes (e.g., knowing where to look in an equation to obtain the slope). Practice in mapping across representations requires accurate selection of information in each representational type and may also lead to intuitions about the way equivalent structures relate across representational types (e.g., learning the graphical consequences of slopes <1 or negative intercepts). All of these notions have been explicitly instructed earlier in the mathematics curriculum; moreover, the PLM contained no additional explicit instruction (other than feedback indicating the correct answer on each trial). It was predicted that improvements in selective information extraction and mapping in the

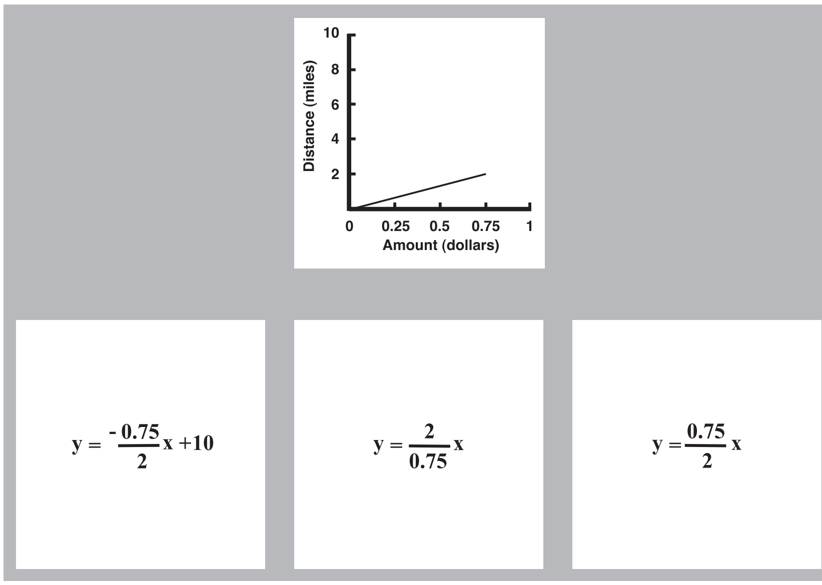

Fig. 1. Example of a display in the MultiRep PLM.

PLM would transfer to important core mathematical tasks, such as generating a correct equation from a word problem or a correct graph from an equation.

7.1. Method

7.1.1. Participants

Sixty-eight ninth and tenth grade students, taking algebra or geometry at a diverse private school in Santa Monica, California, participated in this study.

7.1.2. Design

Students received a paper-and-pencil pretest and posttest containing two kinds of problems. Four problems required solving word problems involving linear functions. Eight *translation* problems involved presentation of a word problem, graph, or equation with the student being asked to translate the given target to a new representation—specifically, to generate an appropriate graph or equation in response. There were four types of translation problem: equation to graph (EG), graph to equation (GE), word problem to equation (WE), and word problem to graph (WG). Students were not asked to generate word problems (i.e., equation or graph to word problem) because of the variability in possible correct responses.

Students in the PLM condition used a self-contained computer program that ran on a Windows platform with a point-and-click interface. The PLM consisted of short mapping trials, where students were presented with a target equation, graph, or word problem, and were asked to select among three possible choices a representation depicting the same information (Fig. 1). Mapping trials used all possible pairs of word, equation, and graphical representations given as targets and choices (with the constraint that the target and choices were in different formats). All equations were in the slope-intercept (y = mx + b) form. The program tracked responses and speed. Visual and auditory feedback indicated whether each student response was correct, and if not, what the correct answer was. Training consisted of two sessions of 60 trials each.

In a control condition, students were asked to practice the same kinds of translation problems that appeared on the assessments. They were given packets with 32 problems including equal numbers of the four generation problem types, designed to closely resemble the translation problems on the assessments. Every time students completed a section of the practice packet, they were given an answer key to check their answers. Feedback stated the correct answer and offered no further explanations. For both the control learning condition, and the assessments in both conditions, paper-and-pencil tests were used to give students flexibility in generating graphs and equations. Time on task for the control was matched to the average time required by participants in the PLM condition.

7.1.3. Procedure

The students used two class periods on two consecutive days to complete the pretest, the instructional intervention, and the posttest. On the first day, students completed a brief background questionnaire, the pretest, and began their learning intervention (either practice packets or PLM). On the second day, students completed their learning interventions and took the posttest. Control and PLM conditions took comparable amounts of time.

7.2. Results and discussion

Primary results for translation problems for the PLM and control conditions are shown in Fig. 2. There were no significant differences in pretest accuracy between the control and PLM conditions, $t(67) = 1.11$, $p = .27$. There was a robust interaction of test by condition, $F(1,66) = 21.17$, indicating that the PLM group improved from pretest to posttest more than the control group. Word problem solving (not shown) averaged about 85% in the pretest in both groups and did not vary between groups.

These results, from two short sessions of PLM use, indicate that practice in mapping problems across multiple representations led to strong improvements on a transfer task—generating the correct equation or graph from a word problem, graph, or equation. In contrast, for the Control group, the translation task in the posttest was not one of transfer; it was the same task practiced during training. The remarkable fact that the PLM group performed better on this transfer task than a control group that practiced the actual task suggests that the PLM produced improvements in structure extraction that are useful to other mathematical skills, such as representation generation. We have also studied this PLM with other age groups. For comparison, we include here a 12th grade sample (Fig. 1, rightmost data). Pretest scores indicate that even in grade 12, the initial ability to generate correct equations and graphs is poor.

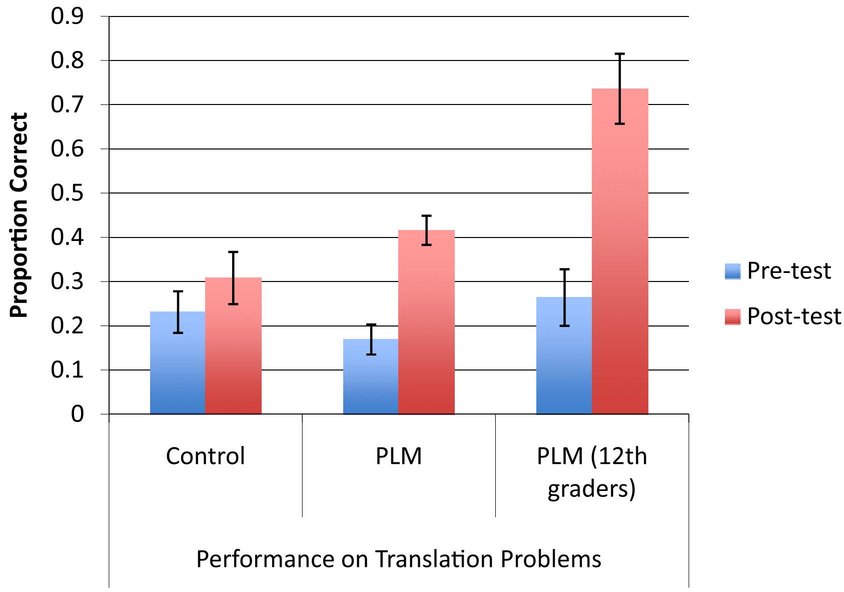

Fig. 2. Results of MultiRep PLM study for translation problems by condition and test. Translation problems required students to generate a correct equation or graph given a word problem, graph, or equation. Control and PLM conditions from the experiment are shown in the left and middle sections. A comparison group of 12th graders run in a separate sample is shown in the rightmost column. Error bars indicate ±1 standard error of the mean.

The current results do not provide insight regarding every variable that differed between the experimental and control groups. In some sense, both manipulations encouraged mapping structure across representations; it appears that the organized trials of the PLM were more effective. One important difference may have been the fact that the PLM group received feedback after each trial. Not only might this have facilitated PL, but immediate feedback can be valuable in various learning contexts (e.g., Mathan & Koedinger, 2003). We have carried out further studies attempting to isolate effective ingredients of the PLM; these raise a number of interesting issues and will be reported elsewhere (J. Son, J. Zucker, N. Chang, & P. J. Kellman, unpublished data).

The MultiRep PLM included the three design properties described earlier. Students performed discriminations and mappings involving key structures in the PLM task; they did not receive explicit procedural or declarative instruction, and they did not solve equations or word problems in the PL task. The results also reflect outcomes that we suggested are consistent with PL effects. The core notions in this module (e.g., the structure of the equations given in slope-intercept form, the depiction of linear functions in Cartesian coordinates, and the interpretation of word problems) are heavily instructed topics in middle and high school mathematics curricula. The improvement produced by a short PLM intervention at both grade levels is generally consistent with the idea that the PLM addressed dimensions of learning that are not well-addressed by conventional instruction. The results also indicated generative use of structure. Achieving learning criteria within the PLM required accurate and fluent processing of novel exemplars. Moreover, evidence of remote transfer was found, in learners' markedly enhanced abilities to generate correct equations and graphs from word problems, a correct equation from a graph, and so on.

8. Experiment 2—From knowledge to fluency: Algebraic transformations

One prediction of a PL approach is that it should be possible for a student to have relevant declarative and procedural knowledge in some domain and yet lack fluent information extraction skills. We tested this idea in work in algebra learning with students who had been instructed for half of a school year on the basic concepts and procedures for solving equations. The hypothesis was that despite reasonable student success in declarative and procedural learning, the ''seeing'' part of algebra is poorly addressed by ordinary methods and might be accelerated by a PL intervention focused on structures and transformations.

The task we chose was mapping algebraic transformations. On each trial, a target equation appeared, and below it was given four other equations. One equation was a legal algebraic transformation of the target; the others were not. The learner was instructed to choose the legal transform as accurately and quickly as possible. The task was constructed to require comparison of structure between the target equation and possible choices. Targets and choices were novel on each trial, and pretest and posttest problems were not used in the learning phase. The PLM design incorporated the design criteria described above. In particular, learners did not practice solving equations in the PLM. We hypothesized, however, that this PL task, by inducing attention to structure and transformation, would improve the seeing of patterns and relations in algebra, and perhaps transfer to improved fluency in actual problem solving. Additional discussion may be found in Kellman et al. (2008).

8.1. Method

8.1.1. Participants

Participants were 30 eighth and ninth grade students at an independent philanthropic school system in Santa Monica, California, tested after mid-year of a year-long Algebra I course.

8.1.2. Apparatus and materials

The PLM was tested on standard PCs using the Windows operating system in computer-equipped classrooms. All assessments and the PLM were presented on computer, with participants' data being sent to a central server.

8.1.3. Design and procedure

The experiment was set up to assess the effects of PL techniques on learners' speed and accuracy in recognizing algebraic transformations and the transfer of PL improvements in information extraction to algebra problem solving. A pretest was given on one day, followed by 2 days in which students worked on the PLM for 40 min/day. A posttest was administered the next day. For a subset of subjects, a delayed posttest was administered 2 weeks later. In the *Algebraic Transformations 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 Fig. 3. Problems involved shifts of constants, variables, or expressions. Accuracy and speed were measured, and feedback was given. Parallel versions of assessments were constructed such that corresponding problems on separate versions varied in the specific constants, variables, or expressions appearing in each equation. Each participant saw a different version in pretest, posttest, and delayed posttest, with order counterbalanced across participants. Each version of the assessment contained *recognition* problems similar to those in the PLM and *solve* problems, used as a transfer test. Solve problems were basic Algebra I equations in a single variable, ranging from simple items (such as $x - 5 = 2$) to more complex ''two-step'' problems (such as $-6 = 3t/5$).

8.2. Results and discussion

Fig. 4 shows the data from this study on the transfer task of equation solving, for students who completed the pretest, learning phase, posttest, and delayed posttest. A key insight from this study comes from the pretest data. The accuracy of algebra problem solving was quite high for learners at the beginning of the study, averaging almost 80%. This level of competence indicates the success of instructional efforts in conveying concepts and procedures for solving equations. Yet students' explicit knowledge contrasts with an obvious difficulty in fluent processing of structure: Students take about 28 s per problem to solve simple algebra problems! This aspect of their problem solving was dramatically improved by PLM use.

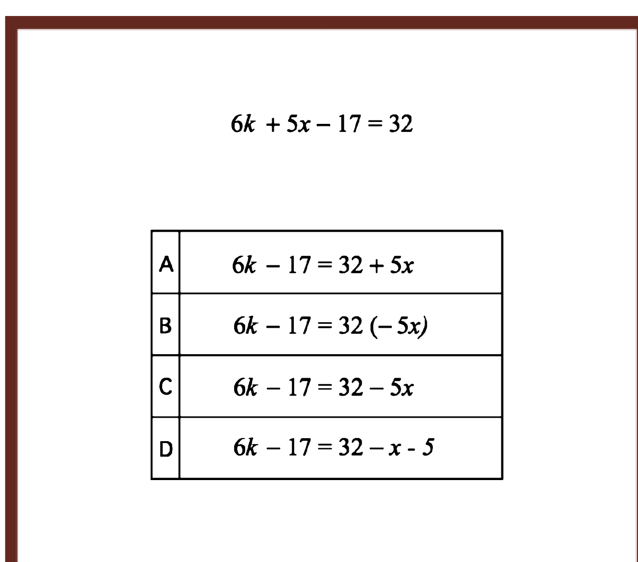

Fig. 3. Example of a display in the Algebraic Transformations PLM.

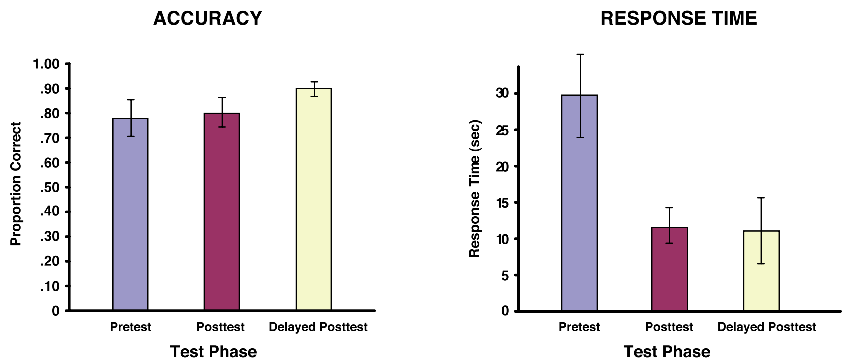

Fig. 4. Results of Algebraic Transformations PLM study for the transfer task of solving algebraic equations. Data for pretest, posttest, and delayed posttest are shown for accuracy (left panel) and response time (right panel). Error bars indicate ±1 standard error of the mean.

After two sessions, speed of solving had dropped to about 12 s per problem. These gains were fully preserved after a 2-week delay. Note that these learners never practiced solving equations in the learning phase; the PLM activity focused on recognizing structure and transformations. These outcomes are consistent with several of the characteristic PL effects we noted earlier: increased fluency, generative use of structure, and persistence over a delay. Perhaps most striking, the attainment of large and lasting gains in fluency from a short intervention suggests that PL methods can produce rapid advancement on dimensions of learning that are not well addressed by conventional instruction.

9. Experiment 3—Fostering structural insight: Linear measurement

U.S. students perform poorly on measurement problems on national and international standardized tests. Even basic skills, such as linear measurement with rulers, show significant deficits. For example, National Assessment of Educational Progress (NAEP) results indicate that many elementary and middle school students are unable to use a ruler that has been broken to measure a 2½ inch toothpick whose left end is aligned with 8 rather than 0 (National Center for Education Statistics, 2008). Students' incorrect responses suggest that they do not conceive of units of linear measurement as having *extent*. Further, they do not make a clear distinction between position and distance, and they have great difficulty using fractions to represent subdivisions of units.

In a current project, we are applying PL principles to concepts of measurement and fractions. One PLM addressed learning difficulties related to linear measurement. Explanations and demonstrations that students normally receive may be insufficient for them to extract relevant features and relations in measurement. As a result, they learn blind procedures that involve misunderstandings of measurement. We applied PL principles using interactive trials that emphasized students' discrimination of position and distance and fostered their structural intuitions about units, including fractional units, in measurement problems.

9.1. Method

9.1.1. Participants

Participants were 63 sixth graders who participated in a PLM instructional intervention plus 78 seventh graders and eighth graders who served as uninstructed control participants, all from the same urban public middle school serving a predominantly low-income neighborhood. Control group participants were highly similar in terms of socioeconomic status, race, and gender to the treatment group (both groups included about 30% African American students, 56% Latino students, 7.5% Asian students, and 4% Caucasian students). All had used the same sixth grade curriculum, and because of standardization in the school district instituted in 2003, the groups received the same curricular units taught in the same order. Many students in the control group had had the same sixth grade teachers as participants in the treatment group.

9.1.2. Design

The Web-delivered PLM presented learners with a graphic display showing a ball on top of a ruler and a billiard cue poised to strike it. Learners were presented with four types of trials that varied the information given and what information was to be found (e.g., given the start and endpoint, find the distance traveled; or given the start point and distance traveled, find the endpoint). The user entered responses by keying them in using an onscreen interface or by dragging a marker on the ruler to the desired point. Once the learner had entered his or her response and pressed a button labeled ''strike,'' the billiard cue would carry out the event on the screen. Animated feedback was provided on each trial.

The learning items in the database varied in numerical values, whether rulers were fully or partially labeled, and whether they were partitioned in the most economical way to solve the problem or were over-partitioned (e.g., a ruler marked in units of 1/16 for a problem involving 1/8s). Items in the learning set were classified into eight categories, including both fraction and integer problems.

9.1.3. Procedure

The sixth grade students first completed a 44-point pencil-and-paper assessment with a variety of items related to linear measurement with integers and fractions, and adding and subtracting fractions. Equivalent versions were used in counterbalanced fashion for posttests and delayed posttests. Virtually all items were transfer items in that they did not directly resemble the trials presented to students during the PLM training. The control group of seventh and eighth graders, who did not participate in any study-related instruction, were administered the assessment just once, providing a baseline comparison for the sixth graders' scores. These seventh and eighth grade control participants received substantial instruction in relevant measurement concepts; in particular, the seventh grade curriculum included 3 weeks on measurement, which control students had completed prior to pretesting.

Intervention participants (sixth graders) received all treatment and posttests prior to exposure to this unit.

After completing the pretest, sixth graders participated in a single introductory classroom lesson lasting 40 min that served to introduce the PLM and the main concepts involved. They then used the PLM software until they either met mastery criteria for all categories or until they had completed six sessions. Meeting mastery criteria could occur within 2–6 sessions; the mean number of sessions for the PLM group was 4.06. Within 1–2 days of completing their last PLM session, students completed a posttest. Four and a half months later, the sixth graders completed a delayed posttest, with no study-related activities occurring in the interim.

9.2. Results and discussion

As can be seen in Fig. 5, prior to instruction, the sixth graders and the seventh and eighth grade control groups scored similarly. This result suggests that the substantial focus on these topics in normal curricula for these grades produces little improvement through the middle school years. PLM use produced significant improvement in the sixth grade intervention group, confirmed by a one-way ANOVA comparing the sixth, seventh, and eighth grade groups [$F(2,138) = 19.687$, $p < .001$]. The sixth graders achieved nearly identical scores on a delayed posttest administered 4.5 months later, indicating that their learning gains were fully maintained. Among the subscales, students' performance improved in reading and constructing lengths with conventional and broken rulers, and they also made strong gains in problems involving fractions. As assessment problems were transfer items of varied kinds, it appears that the PLM guided students to see the

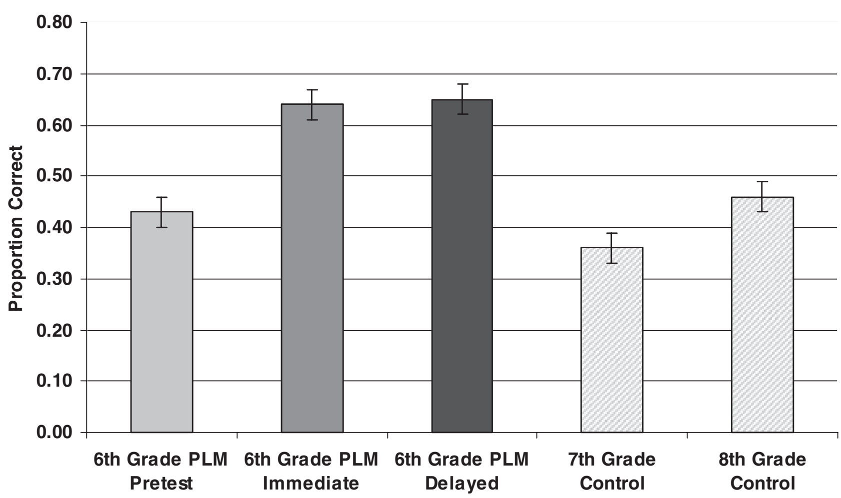

Fig. 5. Results of Linear Measurement PLM. Pretest, posttest, and posttest accuracy after a 4-month delay on a battery of measurement and fraction problems are shown in the leftmost three columns. Seventh and eighth grade control groups are shown in the two columns to the right. Error bars indicate ±1 standard error of the mean.

relevant structures underlying units, measurement, and fractions, replacing blind procedures used initially by many students.

10. General discussion

The study of PL interventions in education and training has barely begun, yet the promise is already clear. PL techniques have the potential to address crucial, neglected dimensions of learning. These include selectivity and fluency in extracting information, discovering important relations, and mapping structure across representations. Each PLM described here addressed an area of mathematics learning known to be problematic for many students. In each case, a relatively short intervention produced major and lasting learning gains, and in each case the learning transferred to key mathematical tasks that differed from the training task.

How do we know that students' learning gains involved PL? As noted earlier, it is difficult in realistic learning domains to exclude all but one type of learning. Both the design and results of the interventions reported here, however, implicate PL as the primary driver of learning. All of the PLMs described here incorporate the general design criteria for PL interventions that we noted. The primary learning tasks required learners to classify or distinguish key features and relations that carry important mathematical information in each domain. Learning occurred over numerous short classification trials with varied instances and involved minimal explicit instruction. Moreover, the results of these interventions show important signatures of PL effects. All of the PLMs showed generativity in structure use, as evidenced by transfer to tasks that differed from the training task. The three PLMs described here also have complementary characteristics with regard to PL effects. The Algebraic Transformations PLM particularly highlights fluency effects in extracting pattern information and the importance of PL manipulations in attaining it. Both the Algebraic Transformations and Linear Measurement PLMs showed no decrement in performance in a delayed posttest. Full preservation of learning gains after a 4.5-month delay in the measurement PLM is an especially striking result. Although persistence of learning may not exclusively implicate PL effects, it is consistent with them.

Perceptual learning methods bear interesting relations to other work applying cognitive principles to improve learning. Some researchers have found that learning of concepts, relations, or problem-solving strategies can be facilitated by comparisons. A number of studies have shown that learning can be enhanced when learners consider two or more cases that involve a common structure but differ in superficial respects (Gick & Holyoak, 1983; Loewenstein, Thompson, & Gentner, 2003); others indicate that comparison of contrasting cases can produce better understanding of relevant structure (Bransford & Schwartz, 2001; Gick & Paterson, 1992; Rittle-Johnson & Star, 2007; Star & Rittle-Johnson, 2009).

Although sometimes described as a cognitive or learning mechanism, ''comparison'' denotes a procedure. It remains to be determined what information processing effects are triggered by comparison, that is, what learning mechanisms are engaged. One answer to this question was suggested by Schwartz and Bransford (2004; see also Bransford & Schwartz, 2001). They refer to:

> ...theories of perceptual learning that emphasize differentiation .... These theories propose that opportunities to analyze sets of contrasting cases .... can help people become sensitive to information that they might miss otherwise .... Contrasting cases help people to notice specific features and dimensions that make the cases distinctive.

This interpretation of contrasting cases is highly consistent with views of PL and E—Gibson's view in particular. Gibson (1969) emphasized differentiation, specifically, the learning of ''distinguishing features,'' in PL. Likewise, the converse idea of comparing varied instances that share some structure describes conditions Gibson noted were relevant to discovery of invariance.

Both of these effects can be modeled in rigorous ways in some PL tasks (e.g., Petrov et al., 2005), although usually in tasks much simpler than those studied here. Still, it has been argued that PL tasks of varying levels and complexity share an underlying commonality in terms of processes of discovery and selection (Kellman & Garrigan, 2009). In neural network approaches to PL, for example, discrimination or classification experience, along with feedback, can lead to the strengthening of the weights of analyzers that detect certain information and downweight other analyzers. Such a mechanism concurrently handles both discovery of distinguishing features between categories and invariance within categories. This description is not meant to oversimplify the modeling task in PL; indeed, high-level PL probably involves generation of candidate structures that are not initially on some fixed list of properties or analyzers, an aspect of discovery that remains mysterious (Kellman & Garrigan, 2009). However, PL models do suggest some promising avenues for understanding the learning mechanisms underlying comparison.

More could be said about comparison procedures and PL interventions; we mention a few interesting issues here. First, whether comparison of a couple or a few cases suffices or whether more extensive classification of examples is needed may depend on the learning task. Second, comparison tasks, more so than simple PL interventions, often involve some explicit instruction. The value of such input, however, may not indicate non-PL factors but simply indicate that language may help guide search and discovery processes in PL (Kellman & Garrigan, 2009). Not much evidence suggests that language alone can accomplish effective discovery; most experiments using comparison appear to rely heavily on representations provided to learners. Furthermore, even when an invariant or distinguishing feature has been explicitly taught, such a manipulation likely does little to enhance fluency of classification.

We note one final issue. Although ''differentiation learning'' has been used as a synonym for PL (Gibson, 1969), it is conceivable that there are domains in which differentiation learning can occur in the complete absence of perceptual representations. Such an effect might allow some uses of comparison to be clearly distinguished from PL; however, as noted above, most comparison experiments make extensive use of representations that permit PL. For example, in a recent study of comparison in numerical estimation procedures (Star & Rittle-Johnson, 2009), learners who compared two estimation strategies viewed 32 worked examples. Clearly, these comparisons provided ample opportunities for PL. Whether there are cases of comparison that do not rely on PL, whether PL and comparison have different effects on fluency, and how explicit inputs may assist discovery processes in PL all pose interesting issues for further research.

Taken together, the results reported here suggest that PL components play a strong role even in complex tasks that involve symbolic processing. They further indicate that appropriately designed PL technology can produce rapid advances in learning. Few learning interventions produce large learning improvements and transfer from short interventions as occurred in each experiment reported here. Further research will undoubtedly reveal even more about how to optimize discovery of structure and fluency in complex domains. Moreover, as we have briefly considered, the synergy in complex tasks among perceptual, declarative, and procedural learning poses important questions and opportunities regarding both the detailed nature of the interactions and how they may be optimally combined in instruction.

Note

1. Although this and some other PLMs utilize a trial format in which several answer choices are presented to the learner, there is no intrinsic connection between multiple choice and PL methods. A number of procedures, including yes/no procedures, 2AFC, adjustment methods, or even free response paradigms, are compatible with PL methods, so long as these tasks require attention to and classification based on differences in pattern or structure. We use a multiple choice format in some PLMs due to its familiarity to students and lower chance accuracy rates than in some other possible response formats.

Acknowledgment

The research reported here was supported in part by grants from the U.S. Department of Education, Institute for Education Sciences, Cognition and Student Learning Program, through grant R305H060070 and from National Science Foundation 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 U.S. Department of Education or the National Science Foundation. We gratefully acknowledge the efforts of Tim Burke, Zipora Roth, Joel Zucker, Amanda Saw, Hye Young Cheong, Warren Longmire, K. P. Thai, Norma Chang, and Joseph Wise. We also wish to thank the students, teachers, and administrators at our partner schools for their roles in making this research possible.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Ahissar, M., Laiwand, R., Kozminsky, G., & Hochstein, S. (1998). Learning pop-out detection: Building representations for conflicting target-distractor relationships. *Vision Research*, *38*(20), 3095–3107.

Baron, J. (1978). The word-superiority effect: Perceptual learning from reading. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes*, Vol. 8. (pp. 131–166). Hillsdale, NJ: Erlbaum.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

Bereiter, C., & Scardamalia, M. (1998). Rethinking learning. In D. Olson & N. Torrance (Eds.), *The handboook of education and human development* (pp. 485–512). Malden, MA: Blackwell.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Bransford, J. D., & Schwartz, D. L. (2001). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*(3), 61–100.

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

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition*, *7*(2), 149–158.

Fahle, M., & Poggio, T. (Eds.) (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: A review. *Journal of Vision*, *2*(2), 190–203.

Garrigan, P., & Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(6), 2248–2253.

Ghose, G. M., Yang, T., & Maunsell, J. H. (2002). Physiological correlates of perceptual learning in monkey v1 and v2. *Journal of Neurophysiology*, *87*(4), 1867–1888.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Prentice-Hall.

Gick, M. L., & Holyaak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology*, *46*, 539–550.

Goldstone, R. L., Landy, D., & Son, J. (2008). A well-grounded education: The role of perception in science and mathematics. In A. Glenberg, M. DeVega, & A. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 327–355). New York: Oxford University Press.

Gómez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135.

Guerlain, S., La Follette, M., Mersch, T. C., Mitchell, B. A., Poole, G. R., Calland, J. F., Jianhong, Lv, & Chekan, E. G. (2004). Improving surgical pattern recognition through repetitive viewing of video clips. *IEEE Transactions on Systems, Man, and Cybernetic – Part A: Systems and Humans*, *34*(6), 699–707.

Kellman, P. J. (2002). Perceptual learning. In H. Pashler & C. R. Gallistel (Eds.), *Stevens' handbook of experimental psychology*, 3rd ed., Vol. 3 (pp. 259–299). New York: John Wiley & Sons.

Kellman, P. J., & Garrigan, P. B. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53–84.

Kellman, P. J., & Kaiser, M. K. (1994). Perceptual learning modules in flight training. *Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society*, *2*, 1183–1187.

Kellman, P. J., Massey, C. M., Roth, Z., Burke, T., Zucker, J., Saw, A., Aguero, K., & Wise, J. (2008). Perceptual learning and the technology of expertise: Studies in fraction learning and algebra. *Learning Technologies and Cognition: Special issue of Pragmatics & Cognition*, *16*(2), 356–405.

Kosslyn, S. M., & Thompson, W. L. (2000). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 975–985). Cambridge, MA: MIT Press.

Kundel, H., & Nodine, C. F. (1975). Interpreting chest radiographs without visual search. *Radiology*, *116*, 527–532.

Landy, D., & Goldstone, R. (2008). How abstract is symbolic thought? *JEP: Learning, Memory and Cognition*, *33*(4), 720–733.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., & Klopfer, D. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hilsdale, NJ: Erlbaum.

Liu, Z. (1999). Perceptual learning in motion discrimination that generalizes across motion directions. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 14085–14087.

Loewenstein, J., Thompson, L., & Gentner, D. (2003). Analogical learning in negotiation teams: Comparing cases promotes learning and transfer. *Academy of Management Learning and Education*, *2*, 119–127.

Massey, C. M., Kellman, P. J., Roth, Z., & Burke, T. (in press). Perceptual learning and adaptive learning technology: Developing new approaches to mathematics learning in the classroom. In N. L. Stein, & S. Rausenbus (Eds.), *Developmental and learning sciences go to school: Implications for education*. New York: Taylor & Francis.

Mathan, S., & Koedinger, K. R. (2003). Recasting the feedback debate: Benefits of tutoring error detection and correction skills. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Artificial intelligence in education, proceedings of AI-ED 2003* (pp. 13–18). Amsterdam: IOS Press.

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

Mollon, J. D., & Danilova, M. (1996). Three remarks on perceptual learning. *Spatial Vision*, *10*(1), 51–58.

National Center for Education Statistics. (2008). *The nation's report card*. Available at: http://nces.ed.gov/nationsreportcard/.

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

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, *6*(3), 189–221.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280.

Rittle-Johnson, B., & Star, J. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*, 561–574.

Robinson, C. S., & Hayes, J. R. (1978). Making inferences about relevance in understanding problems. In R. Revlin & R. E. Mayer (Eds.), *Human reasoning* (pp. 195–206). Washington, DC: Winston.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–8.

Schwartz, D. L., & Bransford, J. D. (2004). A time for telling. *Cognition & Instruction*, *16*(4), 475–522.

Seitz, A., Lefebvre, C., Watanabe, T., & Jolicoeur, P. (2005). Requirement for high-level processing in subliminal learning. *Current Biology*, *15*(18), 753–755.

Seitz, A., & Watanabe, T. (2005). A unified model of task-irrelevant and task-relevant perceptual learning. *Trends in Cognitive Science*, *9*(7), 329–334.

Silva, A. B., & Kellman, P. J. (1999). Perceptual learning in mathematics: The algebra-geometry connection. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty-first annual conference of the Cognitive Science Society* (pp. 683–688). Mahwah, NJ: Erlbaum.

Simon, H. A. (2001). Observations on the sciences of science learning. *Journal of Applied Developmental Psychology*, *21*, 115–121.

Sireteanu, R., & Rettenback, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vision Research*, *40*, 2925–2949.

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: An experimental study on computational estimation. *Journal of Experimental Child Psychology*, *102*, 408–426.

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

Wheeler, D. D. (1970). Processes in the visual recognition of words. *Dissertation Abstracts International*, *31*(2-B), 940.