V International Forum on Teacher Education

Geometric Reformulation of Learning Models Can Help Prepare Better Teachers

Vladik Kreinovich*, Olga Kosheleva

University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA

Abstract

Many researchers have been analyzing how to further improve teacher preparation – and thus, how to improve teaching. Many of their results are based on complex models and/or on complex data analysis. Because of this complexity, future teachers often view the resulting recommendations as *black boxes*, without understanding the motivations for these recommendations – and thus, without much willingness to follow these recommendations. One of the natural ways to make these recommendations clearer is to reformulate them in geometric terms, since geometric models are usually easier to understand than algebraic more abstract ones. In this paper, on the example of two pedagogical recommendations related to the other in which material is represented, we show that the motivations of these recommendations can indeed be described in geometric terms. Hopefully, this will make teachers more willing to follow these recommendations.

Introduction

Teaching is important but difficult, and preparing teachers is also an important and difficult task. Many researchers are analyzing how to improve teacher preparation. When preparing teachers, it is desirable to use the corresponding research results.

―――――――

* Corresponding author. Tel.: +19157576951*; e-mail address*: vladik@utep.edu

Most of these research results are based on complex models and/or on complex data analysis. As a result, future teachers receive the advice as a black box, whose motivation they do not quite understand -- and thus, the advice that they are less inclined to follow than the advice whose motivation was clear. To make sure that the future teacher follow this advice it is therefore desirable to provide motivation that they can reasonably easily understand.

How can we do it? Models usually come in terms of formulas. However, for many future teachers -- just like for many people in general -- a clear geometric picture is easier to understand that the corresponding formula. So, a natural way to make research-based pedagogical recommendations clearer is to reformulate their motivations in geometric form.

In this paper, we provide two examples of such reformulations; specific technical details of these and other examples can be found in Kosheleva (2007), Kosheleva (2010), Kosheleva & Villaverde (2018), and Nguyen & Kreinovich (1997).

Methodology

In this research, we used the general mathematical methodology, ranging from high-school level mathematics (namely, geometry) to more complex invariance and optimization techniques.

Results

General description

In this paper, we provide two cases of the desired reformulation of pedagogy-related mathematical results in geometric terms. Both cases deal with the order in which we present different parts of the material.

The first case deals with a simple dichotomy: shall we start with the material which is the closest to what students know (and which will be easier for them to learn) or with more complex and more challenging material? For example, when learning a new mathematical concept, should we start with easier-to-understand simple examples or should we start with the main ideas of this concept? Intuitively, one would think that starting with the easier material is better, but, somewhat surprisingly, most pedagogical research results show that, in the long run, starting with a more challenging part of the material is more productive

(Kaminsy, Sloutsky, Heckler, 2018; Tchoshanov, 1997). This sounds counterintuitive, but a simple geometric model makes it very convincing (Kosheleva & Villaverde, 2018).

A similar model explains another seemingly counterintuitive result -- that an early start, when children start learning skills much earlier than usual, is often detrimental to their future progress (Kosheleva 2010; Kosheleva & Villaverde, 2018).

The second case explains the advantages of fractal order of presenting the material -- where the optimal fractal dimension depends on the student's learning style (Kosheleva & Villaverde, 2018; Nguyen & Kreinovich, 1997).

Another case for which we have a geometric explanation – but which we did not include here due to lack of space – describes the advantages of a spiral curriculum (Kosheleva, 2007; Kosheleva & Villaverde, 2018), when students repeatedly revisit the same sequence of topics at the increasing level of depth.

Example 1: Early start can inhibit learning -- a geometric explanation

The age at which we teach different topics change. If it turns out that students do not learn, say, reading by the time they should, a natural idea is to start teaching them earlier.

Several decades ago, reading and writing started in the first grade, now they start at kindergarten and even earlier. At first glance, the earlier we start, the better the students will learn.

With the early start, children may play less, their childhood may be not as carefree as it used to be -- but a usual expectation is that with an early start, children will learn better.

In practice, however, early start does not always help: often, early start inhibits learning. For example, according to (Papousek & Papousek, 1977), human infants who started learning to turn their heads to specific sounds at the age of 31 days mastered this task, on average, at the age of 71 days, while infants who started learning this task at birth mastered this task, on average, at the age of 128 days.

This phenomenon is not limited to human infants: according to (Harlow, 1959), an early start in training rhesus monkeys to discriminate objects decreased their peak performance level.

Numerous examples when an early start inhibits learning are presented and discussed in (Bjorklund & Green, 1992), (Bjorklund & Pellegrini, 2002), (Ellis & Bjorklund, 2005), (Bjorklund, 2007), (Remmel, 2008).

The empirical fact that an early start often inhibits learning leads to the following natural question: how to take this phenomenon into account when enhancing student learning?

To be able to take this phenomenon into account in the learning process, we must be able to understand this phenomenon -- and ideally, understand on the quantitative level.

In (Bjorklund & Green, 1992), (Bjorklund & Pellegrini, 2002), (Ellis & Bjorklund, 2005), (Bjorklund, 2007), an attempt is made to understand why early start can inhibit learning. However, the existing understanding is still mostly on the qualitative level, and even on this level, the proposed explanations are still not fully satisfactory; see, e.g., (Remmel, 2008).

The above questions about the efficiency of the early start can be viewed as a particular case of more general questions: what is the best order of presenting the material, the order that leads to the best possible learning?

Many empirical studies have shown that a change in the order in which different parts of the material are presented often drastically changes the learning efficiency; see, e.g., (Van Patten, Chao, & Reigeluth, 1986), (Davydov, 1990), (Tchoshanov, 1997), (Paper & Tchoshanov, 2001), (Lesser & Tchoshanov, 2005), (Lesser & Tchoshanov, 2006), (Kaminski, Sloutsky, & Heckler, 2006), (Kaminski, Sloutsky, & Heckler, 2008).

This is not only about using common sense: sometimes, the empirical results are counter-intuitive. For example: it is usually assumed that most students learn mathematical concepts better if they are first presented with concrete examples of these concepts, and they only learn abstract ideas later on. However, it turns out that empirically, the abstract-first approach for presenting the material often enhances learning; see, e.g., (Tchoshanov, 1997), (Lesser & Tchoshanov, 2005), (Lesser & Tchoshanov, 2006), (Kaminski, Slousky, & Heckler, 2006), (Kaminski, Sloutsky, & Heckler, 2008).

In this section, we attempt to explain the negative effect of early start and, more generally, to explain the reasons why a change in presentation order can drastically change the efficiency of learning. We then show how this explanation can be used to avoid inhibition of learning -- and to enhance the student learning.

To facilitate reasoning about learning, let us start with a simple geometric representation of learning. The process of learning means that we change the state of a student from a state in which the student did not know the material (or does not have the required skill) to a state in which the student has (some) knowledge of the required material (or has the required skill).

Let s_0 denote the original state of a student, and let S denote the set of all the states corresponding to the required knowledge or skill. We start with a state which is not in the set S, and we end up in a state s which is in the set S.

On the set of all possible states, it is natural to define a metric $d(s,s')$ as the difficulty (time, effort, etc.) needed to go from state s to state s'. Our objective is to help the students learn in the easiest (fastest, etc.) way. In terms of the metric d, this means that we want to go from the original state s_0 to some state s from the set S so that the effort $d(s_0,s)$ be the smallest possible.

In geometric terms, the smallest possible effort means the shortest possible distance. Thus, our objective is to find the state s from the set S which is the closest to s_0. Such closest state is called the *projection* of the original state s_0 on the set S.

The above geometric description of learning as a transition from the original state s_0 to its projection on the desired set S describes learning *as a whole*. Our objective is to find out which order of presenting information is the best. Thus, our objective is to analyze the *process* of learning, i.e., learning as a multi-stage phenomenon. For this analysis, we must explicitly take into account that the material to be learned consists of several pieces.

Let S_i, $i = 1, ..., n$, denote the set of states in which a student has learned the i-th part of the material. Our ultimate objective is to make sure that the student learns all the parts of the material. In terms of states, learning the i-th part of the material means belonging to the set S_i. Thus, in terms of states, our objective means that the student should end up in a state which belongs to all the sets $S_1, \ldots, S_n$ -- i.e., in other words, in a state which belongs to the intersectiof the corresponding sets S_i.

In these terms, if we present the material in the order $S_1, \ldots, S_n$, this means that:

- we first project the original state s_0 onto the set S_1, resulting is a state s_1 from S_1 which is the closest to s_0;

- then, we project the state s_1 onto the set S_2, resulting is a state s_2 from S_2 which is the closest to s_1;

- ...

- At the last stage of the cycle, we project the state s_{n-1} onto the set S_n, resulting is a state s_n from S_n which is the closest to s_{n-1};

In some cases, we end up learning all the material -- i.e., in a state s_n that already belongs to the intersection S. However, often, by the time the students have learned S_n, they have somewhat forgotten the material that they learned in the beginning. So, it is necessary to repeat this material again (and again). Thus, starting from the state s_n, we again sequentially project onto the sets $S_1, \ldots, S_n$.

The above «sequential projections» algorithm is actually actively used in many applications; see, e.g., (Gubin, Polyak, & Raik, 1967), (Stark & Yang 1998), (Kontoghiorghes 2006). In the case when all the sets S_i are convex, the resulting Projections on Convex Sets (POCS) method actually guarantees (under certain reasonable conditions) that the corresponding projections converge to a point from the intersection -- i.e., in our terms, that the students will eventually learn all parts of the necessary material.

In the more general non-convex case, the convergence is not always guaranteed -- but the method is still efficiently used, and often converges.

To analyze the problem, let us start with the simplest case when knowledge consists of two parts. In this simplest case, there are only two options. The first option is that we begin by studying S_1, then, we study S_2, then, if needed, we study S_1 again, etc. The second option is that we begin by studying S_2; then, we study S_1, then, if needed, we study S_2 again, etc.

We want to get from the original state s_0 to the state f from the intersection S of S_1 and S_2 which is the closest to s_0. The effectiveness of learning is determined by how close we get to the desired set S in a given number of iterations.

In the case of two-part knowledge, it is natural to conclude that the amount of this knowledge is reasonably small -- otherwise, we would have divided into a larger number of easier-to-learn pieces.

In geometric terms, this means that the original state s_0 is close to the desired intersection set S, i.e., that the distance $d_0 = d(s_0,S)$ is reasonably small.

Since all the states are close to each other, in the vicinity of the final state f, we can therefore expand the formulas describing the borders of the sets S_i into Taylor series and keep only terms which are linear in the (coordinates of the) difference s -- f. Thus, it is reasonable to assume that the border of each of the two sets S_i is described by a linear equation -- and is hence a (hyper-)plane: a line in 2-D space, a plane in 3-D space, etc.

As a result, we arrive at the following configuration. Let $2a$ denote the angle between the borders of the sets S_1 and S_2, so that the angles between each of these borders and the midline is exactly a. Let b denote the angle between the direction from f to s_0 and the midline. In this case, the angle between the border of S_1 and the midline is equal to a -- b.

In the first option, we first project s_0 onto the set S_1. Here, the projection line s_0s_1 is orthogonal to the border of S_1. From the right triangle fs_0s_1, we therefore conclude that the distance $d_1 = d(f,s_1)$ from the projection point s_1 to the desired point f is equal to $d_1 = d_0 * cos(a$ -- $b)$.

On the next step, we project the point s_1 from the set S_1 onto the borderline of S_2 which is located at the angle $2a$ from the borderline of the set S_1. Thus, for the projection result s_2, we will have

$$d_2 = d(s_2,f) = d_1 * cos(2a) = d_0 * cos(a \text{ -- } b) * cos(2a).$$

After this, we may again project onto S_1, then again project onto S_2, etc. For each of these projections, the angle is equal to $2a$, so after each of them, the distance from the desired point f is multiplied the same factor $cos(2a)$. As a result, after k projection steps, we get a point s_k at a distance

$$d_k = d(s_k,f) = d_0 * cos(a \text{ -- } b) * (cos(2a))^{2k \text{ -- } 1}$$

from the desired state f.

In the second option, we start with teaching S_2, i.e., if we first project the state s_k onto the set S_2. In this option, we have $d_1 = d_0 * cos(a + b)$, $d_2 = d(s_2,f) = d_1 * cos(2a) = d_0 * cos(a + b) * cos(2a)$, and, in general,

$$d_k = d(s_k, f) = d_0 * cos(a + b) * (cos(2a))^{2k-1}.$$

Since, in general, *cos(a -- b)* is different from *cos(a + b)*, we can see that a change in the presentation order can indeed drastically change the success of the learning procedure. Thus, our simple geometric model explains why the effectiveness of learning depends on the order in which the material is presented.

Let us extract more specific recommendations from our model. According to the above formulas, starting with S_1 leads to a more effective learning than starting with S_2 if and only if *cos(a -- b) < cos(a + b)*. Since for the angles *x* between 0 and 180 degrees, the cosine *cos(x)* is a decreasing function, we conclude that projection on S_1 is better if and only if *a -- b > a + b,* i.e., if and only if *b < 0.*

Thus, we arrive at the following recommendation: to make learning more efficient, we should start with studying the material which is further away from the current state of knowledge. In other words, we should start with a material that we know the least.

This ties in nicely with a natural commonsense recommendation that to perfect oneself, one should concentrate on one's deficiencies.

This recommendation is also in a very good accordance: with the seemingly counter-intuitive conclusion from (Tchoshanov, 1997), (Lesser & Tchoshanov, 2006), (Kaminski, Sloutsky, Heckler, 2006), (Kaminski, Sloutsky, Heckler, 2008), that studying more difficult (abstract) ideas first enhances learning, and with the human infant studies (Papousek & Papousek, 1977) according to which a concentration on teaching, to human infants, skills that they can easily learn is detrimental in the long run.

Example 2: In what order to present the material -- fractal approach

It is well known that, to achieve good learning, we need to repeat each of the items that a student has to learn. A natural question is: in what order should we present these repetitions? Should we first present all the repetitions of item 1, then all the repetitions of item 2, etc., or should we randomly mix these repetitions?

Let us formulate this problem in precise mathematical terms. Each item is characterized by several (*n*) numerical characteristics, so we can geometrically represent each item as a point in the corresponding *n*-dimensional space.

- Similar items have close values of these characteristics, so the distance between the points corresponding to similar items is small.
- Vice versa, when the items are different, then at least some of these characteristics have different values on these items, so the resulting distance is large.

Thus, the distance between the corresponding points in a multi-D space can be viewed as a measure of similarity between the items.

In terms of multi-D space, an order in which we present repetitions is described as a function $x(t)$, where x is a multi-D point corresponding to the item presented at moment $t = k * d_t$, where d_t is the time between repetitions.

When we have a few items to learn, we can easily learn them all, so there is no need for sophisticated optimization. Optimization becomes necessary when there are many items -- and thus, many repetitions. In this case, similar to the way we simplify the physical problems if we approximate a collection of atoms by a continuous medium, we can approximate the discrete dependence $x(t)$ on discrete time t by a continuous function $x(t)$ of continuous time t.

What is the optimal trajectory $x(t)$? The experience of learning shows that often, presenting the items in random order is beneficial. To allow for this possibility, instead of looking for a deterministic function $x(t)$, we look for *random* processes $x(t)$. Since a deterministic function is a particular case of a random process, we are thus not restricting ourselves.

Let us consider Gaussian random processes. A Gaussian random process can be uniquely characterized by its mean $m(t) = E[x(t)]$ and autocorrelation function $A(t,s) = E[(x(t) -- x(s))^2]$.

Students come with different levels of preparation. Therefore, a good learning strategy should work not only for a student that comes from 0, but also for a student that comes at moment t_0 with the knowledge that other students have already acquired by this time. From this viewpoint, a student's education starts at the moment t_0. It is therefore natural to require that the random process should look the same whether we start with a point $t = 0$ or with some later point t_0. Hence, the characteristics of the process should be the same, i.e., $m(t) = m(t + t_0)$ and $A(t,s) = A(t + t_0, s + t_0)$ for every t, s, and t_0.

From the first condition, we conclude that $m(t) = const$. Thus, by changing the origin of the coordinate system, we can safely assume that $m(t) = 0$.

From the second condition, for $t_0 = --s$, we conclude that $A(t,s) = A(t -- s,0)$, i.e., that the autocorrelation function depends only on the difference between the times: $A(t,s) = a(t -- s)$, where we denoted $a(t) = A(t,0)$. In other words, the random process must be *stationary*.

What autocorrelation function $a(t)$ should we use? The function $a(t)$ depends on how intensely we train. In more intensive training, we present the material faster, and thus, within the same time interval t, we can cover more diverse topics. More diverse topics means that the average change $a(t)$ can be larger. A natural way to describe this increase is by proportionally enlarging all the distances, which leads from $a(t)$ to $C * a(t)$. In other words, if $a(t)$ is a reasonable function for some training, then a new function $C * a(t)$ should also be reasonable.

We can say that the functions $a(t)$ and $C * a(t)$ describe exactly the same learning strategy, but with different intensities. Since intensity can be different, we cannot select a unique function $a(t)$ and claim it to be the best, because for every function $a(t)$, the function $C * a(t)$ describes exactly the same learning strategy. In view of this, instead of formulating a problem of choosing the best autocorrelation *function*, it is more natural to formulate a problem of choosing the best *family $\{C * a(t)\}_C$* of autocorrelation functions.

Among all the families $\{C * a(t)\}_C$, we want to choose the best one. How can we describe what is the best? In mathematical optimization problems, numerical criteria are most frequently used, when to every alternative (in our case, to each family) we assign some value expressing its performance, and we choose an alternative (in our case, a family) for which this value is the largest. In our problem, as such a numerical criterion, we can select, e.g., the average grade A on some standardized test.

However, it is not necessary to restrict ourselves to such numerical criteria only. For example, if we have several different families that have the same average grade A, we can choose between them the one that has the minimal level of uncomfortableness U. In this case, the actual criterion that we use to compare two families is not numerical, but more complicated: *a family F_1 is better than the family F_2 if and only if either $A(F_1) < A(F_2)$, or $A(F_1) = A(F_2)$ and $U(F_1) < U(F_2)$.* A criterion can be even more complicated. What a criterion *must* do is to allow us, for every pair of families, to tell whether the first family is better with respect to this criterion (we'll denote it by $F_1 > F_2$), or the second is better ($F_1 < F_2$), or these families have the same quality in the sense of this criterion (we'll denote it by $F_1 \sim F_2$). Of course, it is necessary to demand that these choices be consistent, e.g., if $F_1 < F_2$ and $F_2 < F_3$ then $F_1 < F_3$.

Another natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family). The reason for this demand is very simple. If a criterion does not choose a family at all, then it is of no use. If several different families are «the best» according to this criterion, then we still have a problem to choose among those «best». Therefore, we need some additional criterion for that choice. For example, if several families turn out to have the same average grade, we can choose among them a with the minimal uncomfortableness.

So what we actually do in this case is abandon that criterion for which there were several «best» families, and consider a new composite criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion or according to the old criterion they had the same quality and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family, it means that this criterion is not final. We have to modify it until we come to a final criterion that will have that property.

The next natural condition that the criterion must satisfy is connected with the fact that the numerical value of the time t depends on the choice of the unit for measuring time. If we replace the original unit of time by a new unit which is m times larger (i.e., replace minutes by hours), the numerical values change from t to t / m. The autocorrelation function that in the old units is described by a family $\{C * a(t)\}$, in the new units, has a new form $\{C * a(m * t)\}$,

Since this change is simply a change in a unit of time, it is reasonable to require that going from $a(t)$ to $a(m * t)$ should not change the *relative* quality of the autocorrelation functions, i.e., if a family $\{C * a(t)\}$ is better that the family $\{C * b(t)\}$, then for every $m > 0$, the family $\{C * a(m * t)\}$ must be still better than the family $\{C * b(m * t)\}$.

Definition 2.1.

- *By an* autocorrelation function *we mean a monotonically non-strictly decreasing function from non-negative real numbers to non-negative real numbers.*
- *By a* family *of functions we mean the family $\{C * a(t)\}_C$, where $a(t)$ is a given autocorrelation function and C runs over arbitrary positive real numbers.*
- *A pair of relations $(<,\sim)$ is called* consistent *if it satisfies the following conditions:*
  (1) if $a < b$ and $b < c$, then $a < c$;
  (2) $a \sim a$;

(3) if a ~ b then b ~ a;

(4) if a ~ b and b ~ c then a ~ c;

(5) if a < b and b ~ c then a < c;

(6) if a ~ b and b < c then a < c;

(7) if a < b, then b < a or a ~ b are impossible.

Definition 2.2.

- *Assume a set A is given. Its elements will be called* alternatives*. By an* optimality criterion*, we mean a consistent pair (<,~) of relations on the set A of all alternatives. If b < a, we say that a is* better *than b; if a ~ b, we say that the alternatives a and b are* equivalent *with respect to this criterion.*

- *We say that an alternative a is* optimal *(or* best*) with respect to a criterion (<,~) if for every other alternative b either b < a or a ~ b.*

- *We say that a criterion is* final *if there exists an optimal alternative, and this optimal alternative is unique.*

- *Let m > 0 be a real number. By the m-*rescaling $R_m(a)$ *of a function a(t) we mean a function a(m * t).*

- *By the m-*rescaling $R_m(F)$ *of a family F, we mean the set of the functions that are obtained from functions a(t) from F by m-rescaling.*

In this section, we consider optimality criteria on the set S of all families.

Definition 2.3. *We say that an optimality criterion on S is* scale-invariant *if for every two families F and G and for every number m > 0, the following two conditions are true:*

- *if F is better than G in the sense of this criterion (i.e., G < F), then $R_m(G) < R_m(F)$;*

- *if F is equivalent G in the sense of this criterion (i.e., G ~ F), then $R_m(G) ~ R_m(F)$.*

As we have already remarked, the demands that the optimalitycriterion is final and scale-invariant are quite reasonable. The only problem with them is that at first glance they may seemrather weak. However, they are not really weak, as the following theorem shows:

Theorem 2.1. (Nguyen and Kreinovich, 1997) *If a family F is optimal in the sense of some optimality criterion which is final and scale-invariant, then every function a(t) from this optimal family F which has*

the form $a(t) = A * t^a$ for some real numbers A and a.

In other words, the optimal configuration is a fractal random process (Mandelbrot, 1983):

- When $a = 2$, we have a straightforward trajectory, without any randomness.
- The value $a = 0$ means that values of $a(t)$ and $a(s)$ for different t and s are completely uncorrelated, i.e., that we have a white noise.
- Intermediate values of a correspond to different levels of fractal randomness.

Our experience showed that such *fractal* order indeed leads to improvement in learning. The exact value of the parameter a -- corresponding to the fractal dimension of the corresponding trajectories -- should be adjusted to the learning style of the students.

Discussions

As we mentioned in the beginning of this paper, geometric explanations of mathematics-based pedagogical recommendations are usually more convincing than the usual algebraic ones, and the move convincing the explanations, the more willingly the teachers will follow them. In this paper, we showed that two important pedagogical recommendations can indeed be reformulated in geometric form – and we also know how to provide a geometric explanation for a third recommendation. We hope that this reformulation will encourage other authors to similarly reformulate other useful mathematical-model-based pedagogical recommendations in geometric form and thus, make teachers more interested in adopting these recommendations.

Conclusion

On the example of two important pedagogical recommendations, we have shown that the motivations for these recommendations can be reformulated in geometric terms, i.e., in terms which are easier to understand. This, hopefully, will make these motivations more convincing and increase the willingness of teachers to follow these recommendations.

References

Bjorklund, D. F. (2007) Why Youth is Not Wasted on the Young: Immaturity in Human Development, Blackwell Publishing.

Bjorklund, D. F. & Green, B. L. (1992) The adaptive nature of cognitive immaturity, *American Psychologist*, 47, 46-54.

Bjorklund, D. F. & Pellegrini, A. (2002) *The Origins of Human Nature: Evolutionary Developmental Psychology*, Washington, DC: American Psychological Association.

Davydov, V. (1990) *Types of Generalizations in Instruction: Logical and Psychological Problems in the Structuring of School Curricular*, Reston, Virginia: NCTM.

Ellis, B. J. & Bjorklund, D. F., eds. (2005) *Origins of the Social Mind: Evolutionary Psychology and Child Development*, New York: Guilford Press.

Gubin, L. G., Polyak, B. T., & Raik, E. V. (1967) The method of projections for finding the comon point of convex sets, *USSR Computational Mathematics and Mathematical Physics*, 7, 1-24.

Harlow, H. W. (1959) The development of learning in the rhesus monkey, *American Scientist*, 47, 459-479.

Kaminski, J. A.. Sloutsky, V. M., & Heckler, A. F. (2016) Do children need concrete instantiations to learn an abstract concept?, In: Sun, R., & Miyake, N. (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, Vancouver, BC, July 26--29, 2006, Mahwah, New Jersey, Lawrence Erlbaum, 411-416.

Kaminski, J. A., Sloutsky, V. M. & Heckler, A .F. (2008) The advantage of abstract examples in learning math, *Science*, 320, 454-455.

Kontoghiorghes, E. J. (2006) *Handbook of Parallel Computing and Statistics*, Boca Raton, Florida: CRC Press.

Kosheleva, O. M. (2007) Spiral curriculum: towards mathematical foundations, *Proceedings of the International Conference on Information Technology InTech'07*, Sydney, Australia, December 12-14, 2007, 152-157.

Kosheleva, O. M. (2010) Early start can inhibit learning: a geometric explanation, *Geombinatorics*, 19(3), pp. 108-118.

Kosheleva, O. M. & Villaverde, K. (2018) *How Interval and Fuzzy Techniques Can Improve Teaching*, Berlin: Springer Verlag.

Lesser, L. M. & Tchoshanov, M. A. (2005) The effect of representation and representational sequence on students' understanding, In: Lloyd, H. M., Wilson, M., Wilkins, J. M. L. & Behm, S. L. (eds.), *Proceedings of the 27th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education PME-NA'2005*, Roanoke, Virginia, October 20-23, 2005.

Lesser, L. M. & Tchoshanov, M. A. (2006) Selecting representations, *Texas Mathematics Teacher*, 53(2), 20-26.

Mandelbrot, B. (1983) *The Fractal Geometry of Nature*, San Francisco, California: Freeman.

Nguyen, H. T. & Kreinovich, V. (1997) *Applications of Continuous Mathematics to Computer Science*, Dordrecht: Kluwer.

Paper, S. & Tchoshanov, M. A. (2001) The role of representation(s) in developing mathematical understanding, *Theory into Practice*, 40(2), 118-127.

Papousek, H. & Papousek, M. (1977) Mothering and the cognitive head-start: Psychobiological considerations. In Schaffer, H. R. (Ed.), *Studies in Mother-Infant Interaction*, London: Academic Press.

Remmel, E. (2008) The benefits of a long childhood, *American Scientist*, 96(3), 250-251.

Stark, H. & Yang, Y. (1988) Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets and Optics, New York: Wiley.

Tchoshanov, M. A. (1997) *Visual Mathematics*, Kazan: ABAK

Van Patten, J., Chao, C. & Reigeluth, C. (1986) A review of strategies for sequencing and synthesizing instruction, *Review of Educational Research*, 56(4), 437-471.