

What Can Help Teach Mathematics Better? Mathematics Itself

Vladik Kreinovich¹ and Olga Kosheleva²

Departments of ¹Computer Science and ²Teacher Education
University of Texas at El Paso
500 W. University, El Paso TX 79968, USA
vladik@utep.edu, olgak@utep.edu

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)



Page 1 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

1. Introduction

- The famous Baron Munchhausen claimed to have pulled himself (and his horse) out of a swamp.
- In the physical world, this is not possible.
- However in the realm of education, a similar phenomenon happens all the time.
- Mathematical models and methods can (and do) help to decide:
 - who teaches better (researcher vs. practitioner, experience vs. training),
 - who studies (shall we teach complex math to anyone or skip people like Pushkin),
 - how to teach (traditional vs. gaming approach),
 - how much feedback to provide when grading,
 - how to best distribute limited tutoring resources, etc.

2. Who Teaches Better: Researcher vs. Practitioner

- Activity or non-activity in research refers to a person's position with respect to knowledge.
- We can consider the set S of all possible combinations of knowledge.
- So they are on the border B of this set.
- Others are placed in the interior of this set.
- In teaching, we are looking for the best value of teaching success $f(s)$.
- In general, the largest value of a function on a set S is attained:
 - either at an interior local maximum,
 - or at one of points in the border.

Introduction

Who Teaches Better...

Shall We Teach...

How To Teach...

How Much Feedback...

How to Best...

Acknowledgments

Bibliography

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 3 of 28

Go Back

Full Screen

Close

Quit

3. Researcher vs. Practitioner (cont-d)

- For any function $f(s)$:
 - if the region S is small enough,
 - the probability that one of the local maxima is contained in this set is close to 0.
- So, with large degree of confidence, the maximum is attained at one of the border points.
- In the great scheme of things, there are many things we still do not know.
- So, the set S describing current knowledge is small.
- Thus, the largest teaching efficiency is attained at the border of S , i.e., for people who are active in research.

Introduction

Who Teaches Better: ...

Shall We Teach ...

How To Teach ...

How Much Feedback ...

How to Best ...

Acknowledgments

Bibliography

Home Page

Title Page



Page 4 of 28

Go Back

Full Screen

Close

Quit

4. Experience vs. Training

- At first glance, several years of practice should help a person pass the certification test.
- In reality, even after several years of practice, many people are not able to pass the test.
- However, after few weeks of intensive training, most people pass it successfully.
- This sounds counterintuitive:
 - the overall number of problems and situations that a person encounters during several years of practice
 - is much larger than what is encountered during a few weeks of intensive training.
- We show that Zipf's law explains this seemingly counterintuitive phenomenon.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)

◀◀

▶▶

◀

▶

Page 5 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

5. Experience vs. Training (cont-d)

- Zipf's law: if we sort all N situations in decreasing order of frequency f_n , we get $f_n \approx \frac{1}{n \cdot \ln(N)}$.
- A certification exam usually includes an equal share of all type of situations.
- To pass a test, a person must correctly reply to a certain proportion p of these problems, e.g., 70 or 80%.
- A practitioner learns by the experience of solving problems of the same type.
- A person learns how to react to situations of given type after encountering several of them.
- Let's denote the needed number of situations by n_0 .
- Suppose that a practitioner solves, on average, s situations per day.

[Introduction](#)[Who Teaches Better: ...](#)[Shall We Teach ...](#)[How To Teach ...](#)[How Much Feedback ...](#)[How to Best ...](#)[Acknowledgments](#)[Bibliography](#)[Home Page](#)[Title Page](#)[◀◀](#)[▶▶](#)[◀](#)[▶](#)[Page 6 of 28](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

6. Experience vs. Training (cont-d)

- Then, during the period of T days, he or she will encounter $T \cdot s$ problems.
- In particular, he/she encounters $T \cdot s \cdot \frac{1}{\ln(N) \cdot n}$ situations of type n .
- He masters situations for which $T \cdot s \cdot \frac{1}{\ln(N) \cdot n} \geq n_0$,
i.e., $n \leq \frac{T \cdot s}{\ln(N) \cdot n_0}$.
- To pass the certification exam, the practitioner needs to learn all situations up to $p \cdot N$.
- Thus, we must have $p \cdot N \leq \frac{T \cdot s}{\ln(N) \cdot n_0}$, i.e.,

$$T \geq T_{\text{practice}} \stackrel{\text{def}}{=} \frac{p \cdot N \cdot \ln(N) \cdot n_0}{s}.$$

7. Experience vs. Training (cont-d)

- What if a person studies for the certification exam?
- In this case, for at least $p \cdot N$ topics, we need to encounter n_0 situations, the total of $p \cdot N \cdot n_0$.
- Thus, if one solves s problems a day, it takes

$$T_{\text{training}} = \frac{p \cdot N \cdot n_0}{s} \text{ days, so } T_{\text{training}} \ll T_{\text{practice}}.$$

- This explains why intensive training is so efficient.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)

[◀◀](#)

[▶▶](#)

[◀](#)

[▶](#)

Page 8 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

8. Shall We Teach Complex Math to Everyone?

- After the same lesson, the amount of learned material often differs drastically, by a factor of 10.
- Does this mean that people have that different learning abilities? Not really.
- Experiments: among different students, learning abilities differ by no more than a factor of 2.
- This fact have been successfully used in designing innovative teaching techniques.
- But why 2, and not 3 or 1.5?
- Learning abilities are, due to genetics, randomly distributed among the human population.
- A natural way to gauge difference between students is to have a numerical value $d(a, b) > 0$ (*metric*).

Introduction

Who Teaches Better...

Shall We Teach...

How To Teach...

How Much Feedback...

How to Best...

Acknowledgments

Bibliography

Home Page

Title Page



Page 9 of 28

Go Back

Full Screen

Close

Quit

9. Shall We Teach Complex Math (cont-d)

- The number n of people on Earth is in billions; so:
 - if there is an asymptotic property of a random metric,
 - then this property is – with high confidence – satisfied for this n .
- A recent result shows that:
 - in almost all randomly selected metric spaces,
 - all the distances are between the largest value and its half.
- The probability of this property tends to 0 as $n \rightarrow \infty$.
- Thus, different distances differ by no more than a factor of 2 – exactly as we observe in the above experiment.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)

◀◀

▶▶

◀

▶

Page 10 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

10. How To Teach (Traditional vs. Gaming Approach): Case of Cybersecurity

- Usually, we explain the main types of cyber-attacks and how to defend against them.
- Lately, a different approach has been very popular and very successful.
- Instead of lectures, students are divided the class into pairs of sparring mini-teams.
- The teams interchangingly try to attack each other and to defend their team from a partner's attacks.
- In the absence of a thorough coverage of all possible topics, one would expect gaps in knowledge.
- However, there are usually no such gaps; why does this work?

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)

◀◀

▶▶

◀

▶

Page 11 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

11. How To Teach (cont-d)

- Is this approach close to optimal or we can drastically further improve it – and if yes, how?
- Let us therefore describe a simple mathematical model that would capture this approach.
- We want is to make sure that:
 - whatever new situation surfaces,
 - the students should have some experience successfully fighting a similar attack in the past,
 - experience that would help the student fight the new attack as well.
- In mathematics, a natural way to describe similarity is by assuming that there is a some metric $d(a, b)$.
- In these terms, “similar” means that $d(a, b) \leq \varepsilon$ for some small threshold value $\varepsilon > 0$.

Introduction

Who Teaches Better: ...

Shall We Teach ...

How To Teach ...

How Much Feedback ...

How to Best ...

Acknowledgments

Bibliography

Home Page

Title Page



Page 12 of 28

Go Back

Full Screen

Close

Quit

12. How To Teach (cont-d)

- We want students to experience situations s_1, \dots, s_n such that every situation s is ε -close to some s_i .
- Such a set is known as an ε -net.
- How do we compare quality of different teaching schemes? The smaller n , the faster we can train.
- Alternatively:
 - we can fix n – and thus, the training time, and
 - try to find the situations s_1, \dots, s_n that lead to the smallest ε .
- The smallest possible number of elements in an ε -net is called ε -entropy.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)

◀◀

▶▶

◀

▶

Page 13 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

13. How To Teach (cont-d)

- It is known that problem of finding the smallest ε -net is, in general, NP-hard; this means that:
 - unless $P = NP$ (which most computer scientists believe to be impossible),
 - no feasible algorithm is possible that would always find the optimal ε -net.
- Let us reformulate adversarial teaching in these terms.
- The first team starts with some attack situation s_1 .
- Then, the sparring team learns how to defend against this attack.
- So, next time, the attacking team will try to find a new way of attacking that has the most chances of success.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)



Page 14 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

14. How To Teach (cont-d)

- It finds the situation s_2 which is as far away from the original situation s_1 as possible:

$$d(s_2, s_1) = \max_{s \in S} d(s, s_1).$$

- In general, once we have experiences the situations s_1, \dots, s_k , we select the next situation s_{k+1} for which

$$\min(d(s_k, s_1), \dots, d(s_k, s_{k-1})) = \\ \max_{s \in S} (\min(d(s, s_1), \dots, d(s, s_{k-1}))).$$

- We continue until we can find a situation which is ε -different from all the previous ones.
- One can prove that this procedure always stops – and generates an ε -net.

Introduction

Who Teaches Better...

Shall We Teach...

How To Teach...

How Much Feedback...

How to Best...

Acknowledgments

Bibliography

Home Page

Title Page

◀◀

▶▶

◀

▶

Page 15 of 28

Go Back

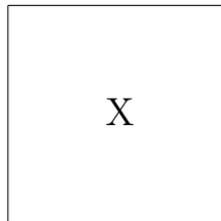
Full Screen

Close

Quit

15. How To Teach (cont-d)

- We can also prove that this strategy is asymptotically optimal:
 - even for one fewer ($n - 1$) situation,
 - the optimal value ε' is at best twice smaller, i.e.,
 $\varepsilon' \geq \varepsilon/2$.
- Graphical illustration: for a unit square, let us pick the midpoint as s_1 :



Introduction

Who Teaches Better: ...

Shall We Teach ...

How To Teach ...

How Much Feedback ...

How to Best ...

Acknowledgments

Bibliography

Home Page

Title Page



Page 16 of 28

Go Back

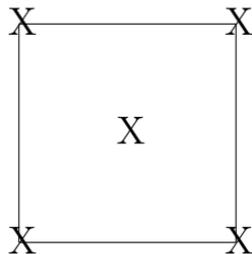
Full Screen

Close

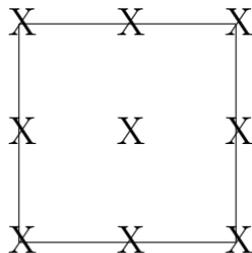
Quit

16. How To Teach (cont-d)

- Then, the next four selections s_i are the vertices:



- After this, the next four selected points s_i are the mid-points of the four edges:



Introduction

Who Teaches Better...

Shall We Teach...

How To Teach...

How Much Feedback...

How to Best...

Acknowledgments

Bibliography

Home Page

Title Page



Page 17 of 28

Go Back

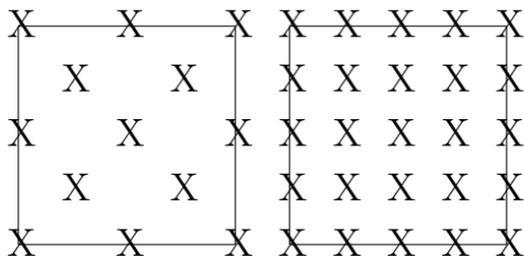
Full Screen

Close

Quit

17. How To Teach (cont-d)

- Here, we have, in effect, four sub-squares.
- On the next stage, the same procedure is repeated for each sub-square, etc.



18. How Much Feedback to Provide?

- In the ideal world, we should assign many homeworks and give a thorough feedback for each homework.
- However, in reality, the instructor's time is limited, so:
 - either we assign few homeworks and give a detailed feed back for all of them,
 - or we assign many homeworks, but give a less thorough feedback.
- What is the optimal thoroughness?
- Let us denote the effect of a piece of feedback that took t hours by $f(t)$.
- Let T denote the overall time allocated to producing this feedback.

Introduction

Who Teaches Better: ...

Shall We Teach ...

How To Teach ...

How Much Feedback ...

How to Best ...

Acknowledgments

Bibliography

Home Page

Title Page



Page 19 of 28

Go Back

Full Screen

Close

Quit

19. How Much Feedback to Provide?

- We want to find the value n and the values t_1, \dots, t_n :
 - for which $t_1 + \dots + t_n = T$, and
 - for which, among all such tuples, the overall effect $E = f(t_1) + \dots + f(t_n)$ is the largest possible.
- Lagrange multiplier technique leads to unconstrained optimization of:

$$f(t_1) + \dots + f(t_n) + \lambda \cdot (t_1 + \dots + t_n - T).$$

- Differentiating w.r.t. each unknown t_i and equating the derivative to 0, we get $f'(t_i) + \lambda = 0$.
- So, all t_i are the same: $t_i = T/n$.
- Thus, we need to maximize the value $E = T \cdot \frac{f(t_1)}{t_1}$.

20. How Much Feedback to Provide?

- Empirically, $f(t) = \frac{t^q}{t^q + c}$, then the optimal time is

$$t_1 = (c \cdot (q - 1))^{1/q}.$$

- Similar formulas can be derived if we take into account that students are different: $f_i(t) \neq f_j(t)$.
- Then, on each assignment to the i -th student, we spend time t_i that maximizes $\frac{f_i(t_i)}{t_i}$.
- Thus, each student gets $(T/n)/t_i$ assignments; so:
 - to some students, we assign fewer homeworks, and provide more thorough feedback, while
 - to other students, we assign more homeworks, and provide less thorough feedback.

Introduction

Who Teaches Better: ...

Shall We Teach ...

How To Teach ...

How Much Feedback ...

How to Best ...

Acknowledgments

Bibliography

Home Page

Title Page



Page 21 of 28

Go Back

Full Screen

Close

Quit

21. How to Best Distribute Tutoring Resources

- In the ideal world, every student who needs tutoring should receive intensive one-on-one tutoring.
- In practice, schools' resources are limited, so the students get only a portion of needed tutoring.
- It would have been not so bad if, e.g., half-time tutoring would be half as efficient as the intensive one.
- However, research shows that half-time math tutoring is, on average, 20 times less efficient.
- To increase the efficiency, we propose to:
 - randomly divide the students who need tutoring into equal-size groups, and
 - each year (or each semester) provide intensive tutoring to only one of these groups.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)



Page 22 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

22. Tutoring (cont-d)

- To further increase the efficiency, we determine, for each student, the optimal number of tutoring per week.
- Suppose that each year, we can devote time T to tutoring for each student who needs tutoring.
- We are considering schools in which education lasts for Y years.
- Thus, during the whole education period, we can allocate tutoring time $T \cdot Y$ to each student.
- If we allocate t hours per week for tutoring, then we can afford to provide this tutoring for $\frac{T \cdot Y}{t}$ weeks.
- Empirically, the weekly grade increase g resulting from time t of tutoring is $g = \frac{g_0}{1 + \exp(-k \cdot (t - a))}$.

Introduction

Who Teaches Better...

Shall We Teach...

How To Teach...

How Much Feedback...

How to Best...

Acknowledgments

Bibliography

Home Page

Title Page



Page 23 of 28

Go Back

Full Screen

Close

Quit

23. Tutoring (cont-d)

- Thus, the overall grade increase based on all this tutoring is:

$$\frac{T \cdot Y}{t} \cdot \frac{g_0}{1 + \exp(-k \cdot (t - a))}.$$

- To find the most efficient arrangement, we need to use the value t that maximizes this expression.
- Differentiating w.r.t. t and equating the derivative to 0, we get the following algorithm for computing this t :
 - first, we find the parameters a and k describing this student's reaction to tutoring;
 - then, we compute the value $C = \exp(k \cdot a)$ and find the solution z of the equation $z - \exp(z) = C$;
 - finally, we compute $t = \frac{z}{a}$.

24. Acknowledgments

This work was supported in part by the National Science Foundation grants:

- 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science);
- HRD-1834620 and HRD-2034030 (CAHSI Includes).

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)

◀◀

▶▶

◀

▶

Page 25 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

25. Bibliography

- Laxman Bokati, Julio Urenda, Olga Kosheleva, and Vladik Kreinovich, “Why immediate repetition is good for short-term learning results but bad for long-term learning: explanation based on decision theory”, In: Martine Ceberio and Vladik Kreinovich (eds.), *How Uncertainty-Related Ideas Can Provide Theoretical Explanation for Empirical Dependencies*, Springer, Cham, Switzerland, to appear.
- Olga Kosheleva, Christian Servin, and Vladik Kreinovich, “Why grade distribution is often multi-modal: an uncertainty-based explanation”, *Proceedings of the World Congress of the International Fuzzy Systems Association and the Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS’2019*, Lafayette, Louisiana, June 18–22, 2019, pp. 106–112.

Introduction

Who Teaches Better: ...

Shall We Teach ...

How To Teach ...

How Much Feedback ...

How to Best ...

Acknowledgments

Bibliography

Home Page

Title Page



Page 26 of 28

Go Back

Full Screen

Close

Quit

26. Bibliography (cont-d)

- O. Kosheleva, C. Servin, and V. Kreinovich, “Under limited resources, lottery-based tutoring is the most efficient”, *Applied Mathematical Sciences*, 2020, Vol. 14, No. 15, pp. 705–710.
- C. Servin, O. Kosheleva, and V. Kreinovich, “Adversarial teaching approach to cybersecurity: a mathematical model explains why it works well”, *Proceedings of the 24th International Conference on Information Visualisation IV’2020*, Vienna and Melbourne, September 7–11, 2020, pp. 313–316.
- C. Servin, O. Kosheleva, and V. Kreinovich, “A recent result about random metrics explains why all of us have similar learning potential”, In: M. Ceberio and V. Kreinovich (eds.), *How Uncertainty-Related Ideas Can Provide Theoretical Explanation for Empirical Dependencies*, Springer, Cham, Switzerland, to appear.

[Introduction](#)

[Who Teaches Better: ...](#)

[Shall We Teach ...](#)

[How To Teach ...](#)

[How Much Feedback ...](#)

[How to Best ...](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)



[Page 27 of 28](#)

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)

27. Bibliography (cont-d)

- Francisco Zapata, Olga Kosheleva, and Vladik Kreinovich, “Being active in research makes a person a better teacher and even helps when working for a company”, *International Mathematical Forum*, 2020, Vol. 15, No. 8, pp. 417–420.
- Francisco Zapata, Olga Kosheleva, and Vladik Kreinovich, “Several years of practice may not be as good as comprehensive training: Zipf’s law explains why”, *Mathematical Structures and Modeling*, 2020, Vol. 54, pp. 145–148.
- Francisco Zapata, Olga Kosheleva, and Vladik Kreinovich, ”Grading homeworks, verifying code: how thorough should the feedback be?”, *Proceedings of the 4th International Conference on Intelligent Decision Science IDS’2020*, Istanbul, Turkey, August 7–8, 2020.

[Introduction](#)

[Who Teaches Better: . . .](#)

[Shall We Teach . . .](#)

[How To Teach . . .](#)

[How Much Feedback . . .](#)

[How to Best . . .](#)

[Acknowledgments](#)

[Bibliography](#)

[Home Page](#)

[Title Page](#)



Page 28 of 28

[Go Back](#)

[Full Screen](#)

[Close](#)

[Quit](#)