Mathematical
Structures and Modeling
2020. N. 2(54). PP. 145–149

UDC 519.2 : [377 + 378]
DOI 10.24147/2222-8772.2020.2.145-149

SEVERAL YEARS OF PRACTICE MAY NOT BE AS GOOD AS COMPREHENSIVE TRAINING: ZIPF'S LAW EXPLAINS WHY

Francisco Zapata
Ph.D. (Phys.-Math.), Instructor, e-mail: fcozpt@outlook.com
Olga Kosheleva
Ph.D. (Phys.-Math.), Associate Professor, e-mail: olgak@utep.edu
Vladik Kreinovich
Ph.D. (Phys.-Math.), Professor, e-mail: vladik@utep.edu

University of Texas at El Paso, El Paso, Texas 79968, USA

Abstract. Many professions practice certifications as a way to establish that a person practicing this profession has reached a certain skills level. At first glance, it may sound like several years of practice should help a person pass the corresponding certification test, but in reality, even after several years of practice, most people are not able to pass the test, while after a few weeks of intensive training, most people pass it successfully. This sounds counterintuitive, since the overall number of problems that a person solves during several years of practice is much larger than the number of problems solved during a few weeks of intensive training. In this paper, we show that Zipf's law explains this seemingly counterintuitive phenomenon.

Keywords: certification, intensive training, practice, Zipf's law.

1. Formulation of the Problem

The more years a person works on a job, the more skilled this person becomes. This is true for medical doctors, this is true for software developers, this is true for college instructors. One may expect that eventually, many years of practical experience would make a person a well-qualified specialist. However, in reality, often people with as many years of practice cannot pass the corresponding qualifying exam (such as a software exam; see, e.g. [1]) — while as little as a few weeks of comprehensive training enables them to pass this exam.

Why is that? It is, to many people, a mystery. It is not that during the previous several years people did not work hard — they did; at times, they make have worked even harder than during their short training, and still, the short training achieves what the previous work did not.

This is not just a theoretical problem. Because of this non-understanding, practitioners with many years of experience often mistakenly think that, because of their experience, they do not need any additional training, and thus miss the opportunity to be successful in getting a certificate for the desired next level of expertise.

The main purpose of this paper is to explain this seemingly mysterious phenomenon and thus, help convince practitioners interested in taking the corresponding qualifying exam to supplement their practical experience with the appropriate comprehensive training.

2. Our Explanation

Main idea behind our explanation: Zipf's law. Our explanation of the above seemingly mysterious phenomenon is based on Zipf's law — an empirical law that describes the frequency of different situation; see, e.g., [2,3]. This law was original discovered in linguistics, where it turns out that if we sort all the words from a natural language in the decreasing order of their frequency, then the frequency f_n of the n-th word is inverse proportional to n: $f_n \approx \dfrac{c}{n}$ for some constant c. A similar dependence holds for many other situations — in particular, for the frequency of different situations, be it programming situations or medical situations.

The constant c can be determined by the fact that when we add up the frequencies of all different words or different situations, we should get 1. Let us describe this idea in precise terms. Let N denote the overall number of possible situations. Then, the above criterion takes the form

$$f_1 + \ldots + f_n + \ldots + f_N = c + \frac{c}{2} + \ldots + \frac{c}{n} + \ldots + \frac{c}{N} = 1,$$

i.e., the form

$$c \cdot \left(1 + \frac{1}{2} + \ldots + \frac{1}{n} + \ldots + \frac{1}{N}\right) = 1.$$

The sum in the right-hand side is an integral sum for the integral $\displaystyle\int_1^N \frac{1}{x}\,dx$, thus it is approximately equal to this integral — which is equal to

$$\ln(N) - \ln(1) = \ln(N) - 0 = \ln(N).$$

Thus, $c \cdot \ln(N) = 1$, so the Zipf's law takes the form

$$f_n = \frac{1}{\ln(N) \cdot n}.$$

So, if we have N types of problems and we sort these types from the most frequent to the least frequent, then the frequency of the problems of n-th type is described by the same Zipf's law formulas.

How we can use this idea to explain the above empirical fact. By definition, a certified professional is a person capable of solving all kinds of problems. Thus, a certification exam usually includes an equal share of all type of problems. To pass a test, a person must successfully solve a certain proportion p of these problems, e.g., 70 or 80%.

How long will it take for a practitioner to gain enough experience to be certified. A practitioner learns by the experience of solving problems of the same type. For simplicity, let us assume that after solving a certain number n_0 of problems of the same type, a person learns to solve *general* problems of this type. A usual number is about $n_0 \approx 10$. Suppose that a practitioner solves, on average, s problems a day. Then, during the period of T days, he or she will encounter $T \cdot s$ problems.

These problems are of different types. For each type n, we can find the overall number of problems of this type if we multiply the overall number $T \cdot s$ of the encountered problems by the frequency f_n of the problems of these types. As a result, we conclude that, on average, the practitioner encountered $T \cdot s \cdot \dfrac{1}{\ln(N) \cdot n}$ problem of the n-th type. So, the practitioners would learn to solve problems of this type if and only if this number is larger than or equal to the learning threshold n_0:

$$T \cdot s \cdot \frac{1}{\ln(N) \cdot n} \geqslant n_0.$$

This is equivalent to

$$n \leqslant \frac{T \cdot s}{\ln(N) \cdot n_0}.$$

Thus, the practitioner learns to solve all the problems of types not exceeding the value $\dfrac{T \cdot s}{\ln(N) \cdot n_0}$.

To pass the certification exam, the practitioner needs to learn to solve all the problems of types up to type $p \cdot N$. Thus, we must have

$$p \cdot N \leqslant \frac{T \cdot s}{\ln(N) \cdot n_0}.$$

This inequality is equivalent to

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

A typical certification exam, whether it is a written exam for a driver's license or a certification exam for C/C++ certification, covers about $N \approx 60$ different topics (to be more precise, subtopics). So, if a person solves, on average, two of three complex problems per day ($s = 2.5$), then, even for the easiest test, with $p = 0.7$, to pass a certification, the person has to work for

$$T_{\text{practice}} = \frac{0.7 \cdot 60 \cdot \ln(60) \cdot 10}{2.5} \approx 840 \text{ days}.$$

Taking into account that every year has 52 weeks with 5 working days a week, i.e., the total of 260 working days, this means that a person who solves two to three complex problems a day would require more than 3 years to become ready.

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

$$T_{\text{training}} = \frac{p \cdot N \cdot n_0}{s}$$

days, i.e., in our case,

$$T_{\text{training}} = \frac{0.7 \cdot 60 \cdot 10}{2.5} \approx 168 \text{ days},$$

i.e., less than half a year. In intensive training, if a person concentrates fully on training, this person can solve four times more complex problems — and thus, finish training four times faster, in about 1.5 months or even less — i.e., in about 6 weeks or so.

If we take into account that many of the attendees of intensive training already have some experience and thus, have already reached the desired skills in several topics, the training time can be further decreased to 5 (or even fewer) weeks.

This is indeed what happens in the actual training — after 5 or 6 weeks of intensive training, most people are able to pass the certifying exam, the same exam that they were not able to pass after two (or even more) years of experience.

Acknowledgments

This work was supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

1. CLA — C Programming Language Certified Associate. URL: `https://cppinstitute.org/cla-exam-syllabus`.
2. Mandelbrot B. The Fractal Geometry of Nature. Freeman, San Francisco, California, 1983.
3. Zipf G.K. Human Behavior and the Principle of Least Effort. Addison-Wesley, 1949.